GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 18:58:05 ; Search time 2378.87 Seconds
(without alignments)
633.372 Million cell updates/sec

Title: US-09-663-600A-139_COPY_36_107
Perfect score: 72
Sequence: 1 atggcctctcttggcctcca.....ttttgggcacactggttgcc 72

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	72	100.0	615	6	AX136475	AX136475 Sequence
2	72	100.0	693	9	AF250558	AF250558 Homo sapi
3	72	100.0	1441	6	AX286822	AX286822 Sequence
4	72	100.0	1475	6	AX092348	AX092348 Sequence
5	72	100.0	1475	6	AX299996	AX299996 Sequence
6	72	100.0	1506	9	BC014424	BC014424 Homo sapi
7	72	100.0	1918	9	AF177340	AF177340 Homo sapi
8	72	100.0	2863	6	AX136129	AX136129 Sequence
9	72	100.0	109465	9	AL158821	AL158821 Human DNA
10	62.4	86.7	953	4	AF358907	AF358907 Canis fam
11	56	77.8	49323	2	AC107455	AC107455 Mus muscu
12	54.4	75.6	779	6	AX286824	AX286824 Sequence
13	54.4	75.6	791	6	E31591	E31591 Tight junct
14	54.4	75.6	791	10	AF072128	AF072128 Mus muscu
15	54.4	75.6	2828	10	BC015252	BC015252 Mus muscu
c 16	39.4	54.7	51	6	AX165471	AX165471 Sequence
c 17	29.6	41.1	198656	2	AC091323	AC091323 Mus muscu
c 18	29.4	40.8	51	6	AX165593	AX165593 Sequence
19	28.8	40.0	720	9	HSA132445	AJ132445 Homo sapi
20	28.8	40.0	1233	9	BC012126	BC012126 Homo sapi
21	28.8	40.0	1938	9	AF314090	AF314090 Homo sapi
22	28.8	40.0	71409	5	AC091535	AC091535 Takifugu
23	28.8	40.0	75015	5	AC091626	AC091626 Danio rer
c 24	28.8	40.0	100058	9	AP000695	AP000695 Homo sapi
c 25	28.8	40.0	101608	9	AP000694	AP000694 Homo sapi
26	28.8	40.0	120000	2	HSAC000005	AC000005 Homo sapi
27	28.8	40.0	175835	2	AC024074	AC024074 Homo sapi
c 28	28.8	40.0	340000	9	AP001726	AP001726 Homo sapi
29	28.6	39.7	1851	9	AF408152	AF408152 Macaca mu
30	28.6	39.7	1869	9	AF408151	AF408151 Macaca mu
31	28.6	39.7	1905	9	AF408153	AF408153 Macaca mu
c 32	28.4	39.4	3643	10	MUSAP1	M61706 Mouse alkal
c 33	28.4	39.4	36901	6	AX015903	AX015903 Sequence
c 34	28.4	39.4	38886	6	AX015909	AX015909 Sequence
c 35	28.4	39.4	228283	10	AF312994	AF312994 Mus muscu
c 36	28	38.9	4046	4	DOGADENCYC	M94968 Canis famil
c 37	28	38.9	4046	6	I29958	I29958 Sequence 1
c 38	28	38.9	173762	2	AC012572	AC012572 Homo sapi
39	28	38.9	199254	2	AC021144	AC021144 Homo sapi
c 40	28	38.9	202032	2	AC087399	AC087399 Homo sapi
c 41	27.8	38.6	728	4	ECU60978	U60978 Equus cabal
c 42	27.6	38.3	90109	2	AC002557	AC002557 Homo sapi
43	27.4	38.1	157041	2	AP004363	AP004363 Oryza sat
44	27.4	38.1	165110	9	AL450311	AL450311 Human DNA
45	27.4	38.1	193219	10	AL606975	AL606975 Mouse DNA

ALIGNMENTS

RESULT 1
AX136475
LOCUS AX136475 615 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 397 from Patent EP1067182.
ACCESSION AX136475
VERSION AX136475.1 GI:14272879
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 615)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
Hayashi,K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 397 10-JAN-2001;
Helix Research Institute (JP)
FEATURES Location/Qualifiers
source 1. .615
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 133 a 174 c 157 g 143 t 8 others
ORIGIN

```
  Query Match            100.0%;  Score 72;  DB 6;  Length 615;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 290

Qy     61 acactggttgcc 72
          ||||||||||||
Db    291 ACACTGGTTGCC 302


RESULT   2
AF250558
LOCUS       AF250558                693 bp    mRNA    linear   PRI 09-AUG-2000
DEFINITION  Homo sapiens claudin-2 mRNA, complete cds.
ACCESSION   AF250558
VERSION     AF250558.1  GI:9755008
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Reinecker,H.-C., Sakaguchi,T. and Golden,H.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Gastrointestinal Unit, Massachusetts
            General Hospital, Fruit Street, Boston, MA 02114, USA
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq22.3-q23"
                     /cell_line="T-84"
                     /tissue_type="epithelium"
     CDS             1. .693
                     /note="tight junction-associated protein"
                     /codon_start=1
                     /product="claudin-2"
                     /protein_id="AAF98151.1"
                     /db_xref="GI:9755009"
                     /translation="MASLGLQLVGYILGLLGLLGLTLVAMLLPSWKTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                     RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                     SEFNSYSLTGYV"
BASE COUNT      142 a    202 c    168 g    181 t
ORIGIN


  Query Match            100.0%;  Score 72;  DB 9;  Length 693;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 60

Qy     61 acactggttgcc 72
          ||||||||||||
Db     61 ACACTGGTTGCC 72


RESULT   3
AX286822
LOCUS       AX286822                1441 bp    DNA     linear   PAT 21-NOV-2001
DEFINITION  Sequence 1 from Patent WO0180879.
ACCESSION   AX286822
VERSION     AX286822.1  GI:17048854
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vinals y de Bassols,C.G., Gaulis,S.R., Coche,T.G. and Orntoft,T.A.
  TITLE     Colorectal cancer vaccines and diagnosis
  JOURNAL   Patent: WO 0180879-A 1 01-NOV-2001;
            SMITHKLINE BEECHAM BIOLOGICALS (S.A.)
FEATURES             Location/Qualifiers
     source          1. .1441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      337 a    399 c    357 g    345 t      3 others
ORIGIN


  Query Match            100.0%;  Score 72;  DB 6;  Length 1441;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 136

Qy     61 acactggttgcc 72
          ||||||||||||
Db    137 ACACTGGTTGCC 148


RESULT   4
AX092348
LOCUS       AX092348                1475 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 79 from Patent WO0116318.
ACCESSION   AX092348
VERSION     AX092348.1  GI:13444488
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1475)
  AUTHORS   Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
            Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
            Wood,W.I.
  TITLE     Secreted and transmembrane polypeptides and nucleic acids encoding
            the same
  JOURNAL   Patent: WO 0116318-A 79 08-MAR-2001;
            Genentech, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1475
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      345 a    404 c    376 g    350 t
ORIGIN


  Query Match            100.0%;  Score 72;  DB 6;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 181

Qy     61 acactggttgcc 72
          ||||||||||||
Db    182 ACACTGGTTGCC 193


RESULT   5
AX299996
```

```
LOCUS       AX299996               1475 bp    DNA     linear   PAT 26-NOV-2001
DEFINITION  Sequence 1 from Patent WO0166740.
ACCESSION   AX299996
VERSION     AX299996.1  GI:17129473
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Eaton,D.L., Fong,S., Goddard,A., Godowski,P.J., Grimaldi,C.J.,
            Gurney,A.L., Tumas,D., Watanabe,C.K., Wood,W.I. and Zhang,Z.
  TITLE     Compositions and methods for the treatment of immune related
            diseases
  JOURNAL   Patent: WO 0166740-A 1 13-SEP-2001;
            Genentech, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1475
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      345 a    404 c    376 g    350 t
ORIGIN


  Query Match             100.0%;  Score 72;  DB 6;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 181

Qy     61 acactggttgcc 72
          ||||||||||||
Db    182 ACACTGGTTGCC 193


RESULT   6
BC014424
LOCUS       BC014424               1506 bp    mRNA    linear   PRI 19-SEP-2001
DEFINITION  Homo sapiens, Similar to claudin 2, clone MGC:20191 IMAGE:4645075,
            mRNA, complete cds.
ACCESSION   BC014424
VERSION     BC014424.1  GI:15680158
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1506)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 29 Row: m Column: 11.
FEATURES             Location/Qualifiers
     source          1. .1506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:20191 IMAGE:4645075"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_15"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             122. .814
                     /codon_start=1
                     /product="Similar to claudin 2"
                     /protein_id="AAH14424.1"
                     /db_xref="GI:15680159"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                     RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                     SEFNSYSLTGYV"
BASE COUNT      377 a    404 c    376 g    349 t
ORIGIN


  Query Match             100.0%;  Score 72;  DB 9;  Length 1506;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 181

Qy     61 acactggttgcc 72
          ||||||||||||
Db    182 ACACTGGTTGCC 193


RESULT   7
AF177340
LOCUS       AF177340               1918 bp    mRNA    linear   PRI 03-OCT-2000
DEFINITION  Homo sapiens clone SP82 claudin 2 mRNA, complete cds.
ACCESSION   AF177340
VERSION     AF177340.1  GI:10503979
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1918)
  AUTHORS   Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
            Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
            Yu,J. and Han,L.H.
  TITLE     Novel human cDNA clone with function of inhibiting cancer cell
            growth
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1918)
  AUTHORS   Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
            Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
            Yu,J. and Han,L.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1999) National Laboratory For Oncogenes & Related
            Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie Tu Road, Shanghai
            200032, P.R. China
FEATURES             Location/Qualifiers
     source          1. .1918
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="SP82"
     CDS             520. .1212
                     /note="similar to Mus musculus claudin 2"
```

```
                     /codon_start=1
                     /product="claudin 2"
                     /protein_id="AAG17984.1"
                     /db_xref="GI:10503980"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                     RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                     SEFNSYSLTGYV"
BASE COUNT      486 a    525 c    473 g    434 t
ORIGIN


  Query Match           100.0%;  Score 72;  DB 9;  Length 1918;
  Best Local Similarity  100.0%;  Pred. No. 4.4e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    520 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 579

Qy     61 acactggttgcc 72
          ||||||||||||
Db    580 ACACTGGTTGCC 591


RESULT   8
AX136129
LOCUS       AX136129               2863 bp    DNA     linear   PAT 30-MAY-2001
DEFINITION  Sequence 51 from Patent EP1067182.
ACCESSION   AX136129
VERSION     AX136129.1  GI:14272537
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2863)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
            Hayashi,K.
  TITLE     Secretory protein or membrane protein
  JOURNAL   Patent: EP 1067182-A 51 10-JAN-2001;
            Helix Research Institute (JP)
FEATURES             Location/Qualifiers
     source          1. .2863
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             231. .923
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC39707.1"
                     /db_xref="GI:14272538"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                     RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                     SEFNSYSLTGYV"
BASE COUNT      670 a    794 c    711 g    688 t
ORIGIN


  Query Match           100.0%;  Score 72;  DB 6;  Length 2863;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 290

Qy     61 acactggttgcc 72
          ||||||||||||
Db    291 ACACTGGTTGCC 302


RESULT   9
AL158821
LOCUS       AL158821               109465 bp    DNA     linear   PRI 08-FEB-2002
DEFINITION  Human DNA sequence from clone RP1-75H8 on chromosome Xq22.3-23
            Contains the CLDN2 gene for claudin 2, the gene for a novel protein
            similar to KIAA0136, a eukaryotic translation elongation factor 1
            alpha 1 (EEF1A1) pseudogene and a CpG island, complete sequence.
ACCESSION   AL158821
VERSION     AL158821.16  GI:14329875
KEYWORDS    HTG; claudin; CLDN2; CpG island; EEF1A1; KIAA0136.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109465)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 8, 2001 this sequence version replaced gi:14160908.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP1-75H8 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone RP1-75H8
            It may be shorter because we sequence overlapping sections only
            once, except for a short overlap.
            The true left end of clone RP1-75H8 is at 1 in this sequence. The
            true left end of clone RP13-383K5 is at 109366 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .109465
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q22.3-23"
                     /clone="RP1-75H8"
                     /clone_lib="RPCI-1"
     repeat_region   1. .62
                     /note="Alu repeat: matches 1. .62 of consensus"
     repeat_region   64. .1090
                     /note="L2 repeat: matches 1154. .2194 of consensus"
     repeat_region   1070. .1202
                     /note="L2 repeat: matches 670. .806 of consensus"
     repeat_region   1558. .1615
                     /note="29 copies 2 mer ct 69% conserved"
     repeat_region   1732. .1858
                     /note="L2 repeat: matches 2579. .2710 of consensus"
     repeat_region   1901. .2047
                     /note="MIR repeat: matches 109. .262 of consensus"
```

```
repeat_region   2112. .2514
                /note="L2 repeat: matches 2095. .2477 of consensus"
repeat_region   2515. .2806
                /note="AluSq repeat: matches 1. .293 of consensus"
repeat_region   2807. .3358
                /note="L2 repeat: matches 1485. .2095 of consensus"
repeat_region   3374. .3580
                /note="AluJb repeat: matches 4. .199 of consensus"
repeat_region   3581. .3610
                /note="15 copies 2 mer ac 100% conserved"
repeat_region   3617. .3684
                /note="FRAM/FAM repeat: matches -4. .63 of consensus"
repeat_region   3685. .3716
                /note="16 copies 2 mer ac 100% conserved"
repeat_region   3717. .3891
                /note="AluJb repeat: matches 127. .301 of consensus"
mRNA            4841. .6387
                /gene="CLDN2"
                /note="match: cDNAs: Em:AF250558 Em:AK004990 Em:AF358908
                Em:AF177340 Em:AB000713 Em:BC014424 Em:BC015252
                Em:AK020255 Em:AF314089 Em:AF087822 Em:AJ011789
                Em:AF358907 Em:AF314090 Em:AF072128 Em:BC012126 Em:D88492
                match: ESTs: Em:AW107793 Em:AA345449 Em:BG164241
                Em:BG423208 Em:T86266 Em:BE304667 Em:BG951731 Em:AA639438
                Em:BE480970 Em:BG951371 Em:BG328625 Em:BG325755
                Em:BG951374 Em:AI790813 Em:AW006504 Em:BF819923
                Em:AI262767 Em:BB847585 Em:BI006997 Em:BG693582
                Em:AI685439 Em:AI949846 Em:BB845010 Em:BG985977 Em:T86158
                Em:BG469035 Em:AA976345 Em:AV604013 Em:AI659323
                Em:BF104997 Em:BB653578 Em:BG385562 Em:AI913916
                Em:BE048787 Em:AW083920 Em:AV647382 Em:AI685788
                Em:BI010035 Em:BI006211 Em:AW837723 Em:AW837724
                Em:AW385836 Em:AW837727 Em:BE513091 Em:BG423109
                Em:BG164062 Em:AI285131 Em:BF541534 Em:AV605969
                Em:AI953481 Em:AI343787 Em:AI769564 Em:AW194732
                Em:BF032123 Em:AW015998 Em:BG426062 Em:AA552647
                Em:BB848306 Em:AI510548 Em:BF758311 Em:BG386074
                Em:BI143586 Em:BG399901 Em:BI008886 Em:AI932964
                Em:AV602536 Em:AV602537"
                /product="dJ75H8.1 (claudin 2)"
                /evidence=not_experimental
gene            4841. .6387
                /gene="CLDN2"
CDS             5019. .5711
                /gene="CLDN2"
                /note="match: proteins: Sw:O88552"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ75H8.1 (claudin 2)"
                /protein_id="CAD23055.1"
                /db_xref="GI:18643740"
                /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                SEFNSYSLTGYV"
polyA_signal    6369. .6374
                /gene="CLDN2"
polyA_site      6387
                /gene="CLDN2"
repeat_region   7257. .7654
                /note="L2 repeat: matches 2322. .2748 of consensus"
repeat_region   7895. .8027
                /note="MIR repeat: matches 76. .201 of consensus"
repeat_region   9628. .9756
                /note="L2 repeat: matches 2566. .2693 of consensus"
repeat_region   10976. .11272
                /note="AluY repeat: matches 1. .299 of consensus"
repeat_region   11462. .11792
                /note="L1PA2 repeat: matches 5821. .6146 of consensus"
repeat_region   12308. .12478
                /note="MIR repeat: matches 2. .179 of consensus"
repeat_region   13066. .13208
                /note="L2 repeat: matches 2359. .2491 of consensus"
repeat_region   14358. .14430
                /note="L2 repeat: matches 2664. .2732 of consensus"
repeat_region   14441. .14555
                /note="L2 repeat: matches 2574. .2692 of consensus"
repeat_region   14511. .14633
                /note="L2 repeat: matches 2570. .2705 of consensus"
repeat_region   14633. .15490
                /note="L2 repeat: matches 1702. .2606 of consensus"
repeat_region   15618. .16068
                /note="L2 repeat: matches 1967. .2418 of consensus"
mRNA            complement(join(17526. .18422,18728. .19011,19305. .19995,
                31703. .31878,33270. .33324,33726. .33794,35126. .35254,
                38802. .38900,53374. .53474,54870. .54989,57681. .57809,
                58135. .58267,61886. .62032,62772. .62990,70020. .70152,
                76467. .76539,76658. .77002))
                /gene="dJ75H8.2"
                /note="The splice sites between exon 4 and 5 are not
                canonical.
                match: cDNAs: Em:AK017472 Em:AK056235 Em:D50926
                Em:AK021627
                match: ESTs: Em:BG962578 Em:BF161943 Em:AW880880
                Em:BG977099 Em:AW881141 Em:AW881142 Em:BB373111
                Em:BB665619 Em:BB649318"
                /product="dJ75H8.2 (similar to KIAA0136 protein)"
                /evidence=not_experimental
gene            complement(17526. .77002)
                /gene="dJ75H8.2"
polyA_site      complement(17526)
                /gene="dJ75H8.2"
polyA_signal    complement(17543. .17548)
                /gene="dJ75H8.2"
repeat_region   18154. .18201
                /note="24 copies 2 mer ga 93% conserved"
CDS             complement(join(18269. .18422,18728. .19011,19305. .19995,
                31703. .31878,33270. .33324,33726. .33794,35126. .35254,
                38802. .38900,53374. .53474,54870. .54989,57681. .57809,
                58135. .58267,61886. .62032,62772. .62990,70020. .70152,
                76467. .76539))
                /gene="dJ75H8.2"
                /note="match: proteins: Sw:Q14149"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ75H8.2 (similar to KIAA0136 protein)"
                /protein_id="CAD23056.1"
                /db_xref="GI:18643741"
                /translation="MSPRYLQSNSSSHTRPFSAIAELLDNAVDPDVSARTVFIDVEEV
                KNKSCLTFTDDGCGMTPHKLHRMLSFGFTDKVIKKSQCPIGVFGNGFKSGSMRLGKDA
                LVFTKNGGTLTVGLLSQTYLECVQAQAVIVPIVPFNQQNKKMIITEDSLPSLEAILNY
                SIFNRENDLLAQFDAIPGKKGTRVLIWNIRRNKNGKSELDFDTDQYDILVSDFDTEEK
                MTGGVTSELPETEYSLRAFCGILYMKPRMKIFLRQKKVTTQMIAKSLANVEYDTYKPT
                FTNKQVRITFGFSCKNSNQFGIMMYHNNRLIKSFEKVGCQVKPTRGEGVGVIGVIECN
                FLKPAYNKQDFEYTKEYRLTINALAQKLNAYWKEKTSQDNFETSTVARPIPKVPDQTW
                VQCDECLKWRKLPGKIDPSMLPARWFCYYNSHPKYRRCSVPEEQELTDEDLCLSKAKK
                QEQTVEEKKKMPMENENHQVFSNPPKILTVQEMAGLNNKTIGYEGIHSPSVLPSGGEE
                SRSPSLQLKPLDSSVLQFSSKYKWILGEEPVEKRRRLQNEMTTPSLDYSMPAPYRRVE
                APVAYPEGENSHDKSSSERSTPPYLFPEYPEASKNTGQNREVSILYPGAKDQRQGSLL

  Query Match             100.0%;  Score 72;  DB 9;  Length 109465;
  Best Local Similarity   100.0%;  Pred. No. 6.7e-14;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5019 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 5078

Qy     61 acactggttgcc 72
          ||||||||||||
Db   5079 ACACTGGTTGCC 5090


RESULT  10
AF358907
```

```
LOCUS       AF358907                953 bp    mRNA    linear   MAM 08-MAY-2001
DEFINITION  Canis familiaris integral membrane protein claudin-2 mRNA, complete
            cds.
ACCESSION   AF358907
VERSION     AF358907.1  GI:13991612
KEYWORDS    .
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 953)
  AUTHORS   Furuse,M., Furuse,K., Sasaki,H. and Tsukita,S.
  TITLE     Conversion of Zonulae Occludentes from Tight to Leaky Strand Type
            by Introducing Claudin-2 into Madin-Darby Canine Kidney I Cells
  JOURNAL   J. Cell Biol. 153 (2), 263-272 (2001)
  MEDLINE   21206012
   PUBMED   11309408
REFERENCE   2  (bases 1 to 953)
  AUTHORS   Furuse,M., Furuse,K., Sasaki,H. and Tsukita,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2001) Cell Biology, Kyoto University, Faculty of
            Medicine, Yoshida-Konoe, Sakyo, Kyoto 606-8501, Japan
FEATURES             Location/Qualifiers
     source          1. .953
                     /organism="Canis familiaris"
                     /db_xref="taxon:9615"
                     /sex="male"
                     /tissue_type="liver"
     CDS             59. .751
                     /note="localizes to tight junctions"
                     /codon_start=1
                     /product="integral membrane protein claudin-2"
                     /protein_id="AAK51433.1"
                     /db_xref="GI:13991613"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWRTSSYVGTSIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIVSVVGM
                     RCTVFCQDSRAKDRLAVVGGVFFIIGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLVAGIILCFSCPLQGNRSDYYDSYQAQPLATRGSPRPGQPPKAK
                     SEFNSYSLTGYV"
BASE COUNT      195 a    289 c    265 g    204 t
ORIGIN


  Query Match           86.7%;  Score 62.4;  DB 4;  Length 953;
  Best Local Similarity  91.7%;  Pred. No. 7.5e-11;
  Matches   66;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggctttgggc 60
           |||||||||| |||||||||||||| |||||||||||||||||| |||||||| ||||||
Db      59 ATGGCCTCTCTCGGCCTCCAACTTGTAGGCTACATCCTAGGCCTCCTGGGGCTGTTGGGC 118

Qy      61 acactggttgcc 72
           || ||||| |||
Db     119 ACCCTGGTGGCC 130


RESULT  11
AC107455
LOCUS       AC107455                49323 bp    DNA     linear   HTG 21-JAN-2002
DEFINITION  Mus musculus clone RP24-68D11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC107455
VERSION     AC107455.1  GI:18252733
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 49323)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-68D11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 49323)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L19885
                Center clone name: 68_D_11
            --------------.
            * NOTE: This record contains 62 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      702: contig of 702 bp in length
            *      703 802: gap of      100 bp
            *      803     1513: contig of 711 bp in length
            *     1514 1613: gap of      100 bp
            *     1614     2279: contig of 666 bp in length
            *     2280 2379: gap of      100 bp
            *     2380     3087: contig of 708 bp in length
            *     3088 3187: gap of      100 bp
            *     3188     3886: contig of 699 bp in length
            *     3887 3986: gap of      100 bp
            *     3987     4687: contig of 701 bp in length
            *     4688 4787: gap of      100 bp
            *     4788     5473: contig of 686 bp in length
            *     5474 5573: gap of      100 bp
            *     5574     6271: contig of 698 bp in length
            *     6272 6371: gap of      100 bp
            *     6372     7070: contig of 699 bp in length
            *     7071 7170: gap of      100 bp
            *     7171     7868: contig of 698 bp in length
            *     7869 7968: gap of      100 bp
            *     7969     8665: contig of 697 bp in length
            *     8666 8765: gap of      100 bp
            *     8766     9473: contig of 708 bp in length
            *     9474 9573: gap of      100 bp
            *     9574    10264: contig of 691 bp in length
            *    10265 10364: gap of      100 bp
```

```
*    10365    11044: contig of 680 bp in length
*    11045 11144: gap of      100 bp
*    11145    11854: contig of 710 bp in length
*    11855 11954: gap of      100 bp
*    11955    12647: contig of 693 bp in length
*    12648 12747: gap of      100 bp
*    12748    13448: contig of 701 bp in length
*    13449 13548: gap of      100 bp
*    13549    14252: contig of 704 bp in length
*    14253 14352: gap of      100 bp
*    14353    15050: contig of 698 bp in length
*    15051 15150: gap of      100 bp
*    15151    15816: contig of 666 bp in length
*    15817 15916: gap of      100 bp
*    15917    16617: contig of 701 bp in length
*    16618 16717: gap of      100 bp
*    16718    17404: contig of 687 bp in length
*    17405 17504: gap of      100 bp
*    17505    18203: contig of 699 bp in length
*    18204 18303: gap of      100 bp
*    18304    18990: contig of 687 bp in length
*    18991 19090: gap of      100 bp
*    19091    19789: contig of 699 bp in length
*    19790 19889: gap of      100 bp
*    19890    20591: contig of 702 bp in length
*    20592 20691: gap of      100 bp
*    20692    21393: contig of 702 bp in length
*    21394 21493: gap of      100 bp
*    21494    22193: contig of 700 bp in length
*    22194 22293: gap of      100 bp
*    22294    22993: contig of 700 bp in length
*    22994 23093: gap of      100 bp
*    23094    23795: contig of 702 bp in length
*    23796 23895: gap of      100 bp
*    23896    24572: contig of 677 bp in length
*    24573 24672: gap of      100 bp
*    24673    25373: contig of 701 bp in length
*    25374 25473: gap of      100 bp
*    25474    26173: contig of 700 bp in length
*    26174 26273: gap of      100 bp
*    26274    26986: contig of 713 bp in length
*    26987 27086: gap of      100 bp
*    27087    27788: contig of 702 bp in length
*    27789 27888: gap of      100 bp
*    27889    28592: contig of 704 bp in length
*    28593 28692: gap of      100 bp
*    28693    29390: contig of 698 bp in length
*    29391 29490: gap of      100 bp
*    29491    30192: contig of 702 bp in length
*    30193 30292: gap of      100 bp
*    30293    30979: contig of 687 bp in length
*    30980 31079: gap of      100 bp
*    31080    31752: contig of 673 bp in length
*    31753 31852: gap of      100 bp
*    31853    32557: contig of 705 bp in length
*    32558 32657: gap of      100 bp
*    32658    33356: contig of 699 bp in length
*    33357 33456: gap of      100 bp
*    33457    34159: contig of 703 bp in length
*    34160 34259: gap of      100 bp
*    34260    34953: contig of 694 bp in length
*    34954 35053: gap of      100 bp
*    35054    35752: contig of 699 bp in length
*    35753 35852: gap of      100 bp
*    35853    36550: contig of 698 bp in length
*    36551 36650: gap of      100 bp
*    36651    37354: contig of 704 bp in length
*    37355 37454: gap of      100 bp
*    37455    38154: contig of 700 bp in length
*    38155 38254: gap of      100 bp
*    38255    38941: contig of 687 bp in length
*    38942 39041: gap of      100 bp
*    39042    39750: contig of 709 bp in length
*    39751 39850: gap of      100 bp
*    39851    40555: contig of 705 bp in length
*    40556 40655: gap of      100 bp
*    40656    41354: contig of 699 bp in length
*    41355 41454: gap of      100 bp
*    41455    42151: contig of 697 bp in length
*    42152 42251: gap of      100 bp
*    42252    42962: contig of 711 bp in length
*    42963 43062: gap of      100 bp
*    43063    43751: contig of 689 bp in length
*    43752 43851: gap of      100 bp
*    43852    44535: contig of 684 bp in length
*    44536 44635: gap of      100 bp
*    44636    45330: contig of 695 bp in length
*    45331 45430: gap of      100 bp
*    45431    46126: contig of 696 bp in length
*    46127 46226: gap of      100 bp
*    46227    46930: contig of 704 bp in length
*    46931 47030: gap of      100 bp
*    47031    47736: contig of 706 bp in length
*    47737 47836: gap of      100 bp
*    47837    48533: contig of 697 bp in length
*    48534 48633: gap of      100 bp
*    48634    49323: contig of 690 bp in length.
FEATURES             Location/Qualifiers
     source          1. .49323
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP24-68D11"
                     /clone_lib="RPCI-24 Male Mouse BAC"
BASE COUNT    12875 a   8561 c   9043 g  12675 t   6169 others
ORIGIN


  Query Match             77.8%;  Score 56;  DB 2;  Length 49323;
  Best Local Similarity   86.1%;  Pred. No. 1.7e-08;
  Matches   62;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||| |||||| ||||||| |||||||||||||||||||| ||||||| || |||
Db  32239 ATGGCCTCCCTTGGCGTCCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 32298

Qy      61 acactggttgcc 72
           |||    |||||
Db  32299 ACATCCATTGCC 32310


RESULT  12
AX286824
LOCUS       AX286824                 779 bp    DNA     linear   PAT 21-NOV-2001
DEFINITION  Sequence 3 from Patent WO0180879.
ACCESSION   AX286824
VERSION     AX286824.1  GI:17048855
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Vinals y de Bassols,C.G., Gaulis,S.R., Coche,T.G. and Orntoft,T.A.
  TITLE     Colorectal cancer vaccines and diagnosis
  JOURNAL   Patent: WO 0180879-A 3 01-NOV-2001;
            SMITHKLINE BEECHAM BIOLOGICALS (S.A.)
FEATURES             Location/Qualifiers
     source          1. .779
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
BASE COUNT      152 a    205 c    216 g    206 t
ORIGIN


  Query Match             75.6%;  Score 54.4;  DB 6;  Length 779;
  Best Local Similarity   84.7%;  Pred. No. 3.8e-08;
```

```
  Matches   61;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db     48 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 107

Qy     61 acactggttgcc 72
          |||    |||||
Db    108 ACATCCATTGCC 119


RESULT  13
E31591
LOCUS       E31591                   791 bp    DNA     linear   PAT 07-FEB-2001
DEFINITION  Tight junction-constituting membrane protein claudin family.
ACCESSION   E31591
VERSION     E31591.1  GI:13018519
KEYWORDS    JP 2000032984-A/2.
SOURCE      Mus sp. .
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Syoichiro,T.
  TITLE     Tight junction-constituting membrane protein claudin family
  JOURNAL   Patent: JP 2000032984-A 2 02-FEB-2000;
            EISAI CO LTD
COMMENT     OS   Mus sp. (mouse)
            PN   JP 2000032984-A/2
            PD   02-FEB-2000
            PF   26-JUN-1998 JP 1998179847
            PR
            PI   SYOICHIRO TSUKITA
            PC   C12N15/09,C07K14/47,C07K16/18,C12N5/10,C12P21/02,C12Q1/02, PC
             C12Q1/68,
            PC   G01N33/15,G01N33/53,G01N33/577//C12P21/08,(C12N5/10,C12R1:91),
            PC   (C12P21/02,C12R1:91),C12N15/00,C12N5/00,(C12N5/00,C12R1:91) CC
              Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             48. .740.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
BASE COUNT      155 a    207 c    220 g    209 t
ORIGIN


  Query Match             75.6%;  Score 54.4;  DB 6;  Length 791;
  Best Local Similarity   84.7%;  Pred. No. 3.8e-08;
  Matches   61;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db     48 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 107

Qy     61 acactggttgcc 72
          |||    |||||
Db    108 ACATCCATTGCC 119


RESULT  14
AF072128
LOCUS       AF072128                 791 bp    mRNA    linear   ROD 21-JUL-1998
DEFINITION  Mus musculus claudin-2 mRNA, complete cds.
ACCESSION   AF072128
VERSION     AF072128.1  GI:3335183
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Furuse,M., Fujita,K., Hiiragi,T., Fujimoto,K. and Tsukita,S.
  TITLE     Claudin-1 and -2: novel integral membrane proteins localizing at
            tight junctions with No sequence similarity to occludin
  JOURNAL   J. Cell Biol. 141 (7), 1539-1550 (1998)
  MEDLINE   98311639
REFERENCE   2  (bases 1 to 791)
  AUTHORS   Furuse,M., Fujita,K., Hiiragi,T., Fujimoto,K. and Tsukita,S.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-1998) Cell Biology, Kyoto University Faculty of
            Medicine, Sakyo-ku, Kyoto 606-8501, Japan
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             48. .740
                     /note="integral membrane protein localizing at tight
                     junctions"
                     /codon_start=1
                     /product="claudin-2"
                     /protein_id="AAC27079.1"
                     /db_xref="GI:3335184"
                     /translation="MASLGVQLVGYILGLLGLLGTSIAMLLPNWRTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAMSSLACIISVVGM
                     RCTVFCQDSRAKDRVAVVGGVFFILGGILGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISALFSLVAGVILCFCSPQGNRTNYYDGYQAQPLATRSSPRSAQQPKAK.
                     SEFNSYSLTGYV"
BASE COUNT      155 a    207 c    220 g    209 t
ORIGIN


  Query Match             75.6%;  Score 54.4;  DB 10;  Length 791;
  Best Local Similarity   84.7%;  Pred. No. 3.8e-08;
  Matches   61;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db     48 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 107

Qy     61 acactggttgcc 72
          |||    |||||
Db    108 ACATCCATTGCC 119


RESULT  15
BC015252
LOCUS       BC015252                 2828 bp    mRNA    linear   ROD 29-OCT-2001
DEFINITION  Mus musculus, claudin 2, clone MGC:18348 IMAGE:4223446, mRNA,
            complete cds.
ACCESSION   BC015252
VERSION     BC015252.1  GI:15929633
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2828)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
```

Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 25 Row: h Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7710003.

```
FEATURES             Location/Qualifiers
     source          1. .2828
                     /organism="Mus musculus"
                     /db_xref="LocusID:12738"
                     /db_xref="taxon:10090"
                     /clone="MGC:18348 IMAGE:4223446"
                     /tissue_type="Kidney, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             126. .818
                     /codon_start=1
                     /product="claudin 2"
                     /protein_id="AAH15252.1"
                     /db_xref="GI:15929634"
                     /translation="MASLGVQLVGYILGLLGLLGTSIAMLLPNWRTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAMSSLACIISVVGM
                     RCTVFCQDSRAKDRVAVVGGVFFILGGILGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISALFSLVAGVILCFSCSPQGNRTNYYDGYQAQPLATRSSPRSAQQPKAK
                     SEFNSYSLTGYV"
BASE COUNT      736 a    697 c    667 g    728 t
ORIGIN
```

```
  Query Match             75.6%;  Score 54.4;  DB 10;  Length 2828;
  Best Local Similarity   84.7%;  Pred. No. 4.4e-08;
  Matches   61;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           |||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db     126 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 185

Qy      61 acactggttgcc 72
           |||    |||||
Db     186 ACATCCATTGCC 197


Search completed: June  2, 2002, 18:58:13
Job time: 11113 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 20:09:02 ; Search time 523.16 Seconds
(without alignments)
236.291 Million cell updates/sec

Title: US-09-663-600A-139_COPY_36_107
Perfect score: 72
Sequence: 1 atggcctctcttggcctcca.....ttttgggcacactggttgcc 72

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	72	100.0	324	20	AAX41060	Human secreted pro
2	72	100.0	467	21	AAC98817	Human pancreatic c
3	72	100.0	530	22	AAD08518	Human secreted pro
4	72	100.0	615	22	AAF93963	Primer specific fo
5	72	100.0	1400	20	AAX97865	Human secreted pro
6	72	100.0	1441	24	AAS16180	Human CASB81 polyn
7	72	100.0	1475	21	AAA37060	Human PRO1356 (UNQ
8	72	100.0	1475	22	AAS15360	cDNA encoding huma
9	72	100.0	1475	22	AAS21489	Human cDNA sequenc
10	72	100.0	1475	22	AAF92097	Human PRO1356 cDNA
11	72	100.0	1475	22	AAF54296	DNA encoding prote
12	72	100.0	1481	22	AAD08497	Human secreted pro
13	72	100.0	1524	20	AAX97818	Human secreted pro
14	72	100.0	1540	22	AAI58013	Human polynucleoti
c 15	72	100.0	1761	22	AAI59799	Human polynucleoti
16	72	100.0	1918	24	ABA04424	Human SP82 protein
17	72	100.0	2742	21	AAA12585	cDNA encoding a me
18	72	100.0	2863	22	AAF93769	Human cDNA encodin
19	71.6	99.4	405	20	AAX41059	Human secreted pro
20	54.4	75.6	779	24	AAS16181	Murine CASB81 poly
21	54.4	75.6	791	21	AAZ89137	Murine clodin 2 cD
c 22	39.4	54.7	51	23	ABL00675	Human silent nonco
c 23	29.4	40.8	51	23	ABL00797	Human amino acid c
24	28.8	40.0	1174	21	AAA37113	Human PRO1571 (UNQ
25	28.8	40.0	1174	22	AAS46101	Human DNA encoding
26	28.8	40.0	1174	22	AAF54431	Probe #47 used in
c 27	28.4	39.4	36901	20	AAZ23892	Murine LOBO genomi
c 28	28.4	39.4	38886	20	AAZ23897	Murine LOBO homolo
c 29	28	38.9	4046	14	AAQ42525	Cardiac adenylyl c
30	27.2	37.8	660	21	AAZ89153	Murine clodin 6 cD
31	27.2	37.8	1231	22	AAF32670	Human cDNA encodin
c 32	26.6	36.9	806	22	AAS35139	Human cardiovascul
c 33	26.6	36.9	818	21	AAC77281	Human ORFX ORF2836
c 34	26.4	36.7	1497	21	AAC40181	Arabidopsis thalia
c 35	26.4	36.7	1515	21	AAC50415	Arabidopsis thalia
c 36	26.2	36.4	1530	23	ABL07695	Drosophila melanog
37	26.2	36.4	3609	23	ABL07694	Drosophila melanog
38	26.2	36.4	6968	23	ABL21538	Drosophila melanog
39	26	36.1	1653	22	AAD21436	ADAM-20dis-Fc fusi
40	26	36.1	1981	21	AAZ50482	Corn sulphate perm
41	26	36.1	2181	20	AAX56461	Human SVPH1-26 pro
42	26	36.1	2426	20	AAX28636	Nucleotide sequenc
43	25.8	35.8	766	23	AAS92167	DNA encoding novel
44	25.8	35.8	783	22	AAH99770	Human protein enco
45	25.8	35.8	783	22	AAI61349	Human polynucleoti

ALIGNMENTS

RESULT 1
AAX41060
ID AAX41060 standard; cDNA; 324 BP.
XX
AC AAX41060;
XX
DT 18-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO: 272.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906554-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01238.
XX
PR 01-AUG-1997; 97US-0905134.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153784/13.

```
DR   P-PSDB; AAY12227.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT   muscle, muscle and heart tissue
XX
PS   Claim 1; Page 402-403; 622pp; English.
XX
CC   AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY01602 and
CC   AAY11994 to AAY12260, respectively. The proteins given represent the
CC   signal peptide and an N-terminal fragment of a secreted protein. The
CC   nucleic acid sequences can be used for producing secreted human gene
CC   products. They can also be used to develop products for diagnosis and
CC   therapy. The proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used
CC   for directing extracellular secretion of a polypeptide or the insertion
CC   of a polypeptide into a membrane, or importing a polypeptide into
CC   a cell.
XX
SQ   Sequence 324 BP; 73 A; 80 C; 94 G; 74 T; 3 other;


  Query Match             100.0%;  Score 72;  DB 20;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 2e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 219

Qy      61 acactggttgcc 72
           ||||||||||||
Db     220 acactggttgcc 231


RESULT   2
AAC98817
ID   AAC98817 standard; cDNA; 467 BP.
XX
AC   AAC98817;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:45.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54052.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 527-528; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 467 BP; 95 A; 132 C; 126 G; 109 T; 5 other;


  Query Match             100.0%;  Score 72;  DB 21;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 136

Qy      61 acactggttgcc 72
           ||||||||||||
Db     137 acactggttgcc 148


RESULT   3
AAD08518
ID   AAD08518 standard; cDNA; 530 BP.
XX
AC   AAD08518;
XX
DT   09-AUG-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 10 cDNA clone HTPIH83, SEQ ID NO:41.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             111..530
FT                   /*tag= a
FT                   /product= "Human secreted protein precursor"
FT                   /transl_except= (pos:516..518, aa:Xaa)
FT                   /note= "Xaa equals any of the naturally occurring
FT                   L-amino acids; CDS does not include stop codon"
FT                   /partial
FT   sig_peptide     111..182
FT                   /*tag= b
FT   mat_peptide     183..530
FT                   /*tag= c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200136432-A2.
XX
PD   25-MAY-2001.
XX
PF   15-NOV-2000; 2000WO-US31162.
XX
PR   19-NOV-1999;   99US-0166415.
PR   30-JUN-2000; 2000US-0215136.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Baker KP,  Young PE;
XX
DR   WPI; 2001-343793/36.
DR   P-PSDB; AAE04228.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition -
XX
PS   Claim 1; Page 425; 509pp; English.
XX
CC   AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC   protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC   AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   18 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin ageing due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 530 BP; 110 A; 145 C; 147 G; 127 T; 1 other;


  Query Match           100.0%;  Score 72;  DB 22;  Length 530;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 170

Qy      61 acactggttgcc 72
           ||||||||||||
Db     171 acactggttgcc 182


RESULT   4
AAF93963
ID   AAF93963 standard; DNA; 615 BP.
XX
AC   AAF93963;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Primer specific for DNA encoding secretory/membrane protein SEQ ID 397.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS   Synthetic.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 4; SEQ ID 397; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 615 BP; 133 A; 174 C; 157 G; 143 T; 8 other;
```

```
  Query Match             100.0%;  Score 72;  DB 22;  Length 615;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 290

Qy     61 acactggttgcc 72
          ||||||||||||
Db    291 acactggttgcc 302


RESULT   5
AAX97865
ID   AAX97865 standard; cDNA; 1400 BP.
XX
AC   AAX97865;
XX
DT   23-SEP-1999  (first entry)
XX
DE   Human secreted protein encoding cDNA #53.
XX
KW   Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW   diagnostic; gene therapy; chromosome mapping; secretion vector; ss.
XX
OS   Homo sapiens.
XX
PN   WO9925825-A2.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-IB01862.
XX
PR   04-SEP-1998;   98US-0099273.
PR   13-NOV-1997;   97US-0066677.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-347472/29.
DR   P-PSDB; AAY36181.
XX
PT   Extended cDNAs encoding secreted proteins
XX
PS   Claim 1; Page 254-255; 307pp; English.
XX
CC   AAX97813-X97906 represent extended cDNA's which encode novel human
CC   secreted proteins (see AAY36129-Y36222) and which have cytostatic,
CC   thrombotic and osteopathic activity. The extended cDNAs can be used to
CC   express secreted proteins or parts of them or to obtain antibodies
CC   capable of binding to the secreted proteins. They may also be used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   Uses also include design of expression vectors and secretion vectors.
XX
SQ   Sequence 1400 BP; 326 A; 392 C; 346 G; 336 T; 0 other;


  Query Match             100.0%;  Score 72;  DB 20;  Length 1400;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     36 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 95

Qy     61 acactggttgcc 72
          ||||||||||||
Db     96 acactggttgcc 107


RESULT   6
AAS16180
ID   AAS16180 standard; cDNA; 1441 BP.
XX
AC   AAS16180;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Human CASB81 polynucleotide.
XX
KW   Human; CASB81; colorectal cancer; colon-associated disease; tumour; ss;
KW   immunogen; cytostatic; antitumour; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..769
FT                   /*tag= a
FT                   /product= "Human CASB81 protein"
FT                   /transl_except= (pos:635..637, aa:Xaa)
FT                   /note= "Xaa= unknown"
XX
PN   WO200180879-A2.
XX
PD   01-NOV-2001.
XX
PF   19-APR-2001; 2001WO-EP04456.
XX
PR   20-APR-2000; 2000GB-0009907.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Vinals De Bassols YC,  Gaulis SRJ,  Coche T,  Orntoft T;
XX
DR   WPI; 2002-041368/05.
DR   P-PSDB; AAU10521.
XX
PT   Use of a claudin 2 polypeptide and polynucleotide as vaccine for
PT   preventing and treating colorectal cancer, and as diagnostic reagents
PT   in context of the cancer -
XX
PS   Claim 8; Page 57; 68pp; English.
XX
CC   The invention relates to the use of CASB81 polypeptides and
CC   polynucleotides for the manufacture of a medicament for treating a
CC   patient suffering from or susceptible to colorectal cancer or other
CC   colon-associated tumours or diseases. The sequences act as agents for
CC   specific prophylactic or therapeutic immunisation against tumours because
CC   they are specifically expressed or highly over-expressed in colorectal
CC   tumours compared to normal cells, and can thus be targeted by
CC   antigen-specific immune mechanisms leading to the destruction of tumour
CC   cells. They can be used in development of an antibody immunospecific for
CC   CASB81 protein, which can be used for diagnosis of tumour cells. Peptides
CC   incorporating epitopes of CASB81 provide suitable immunogens. Presence of
CC   or susceptibility to colorectal cancer can be diagnosed by analysing the
CC   presence or amount of CASB81 DNA expression or CASB81 protein activity.
CC   This sequence represents a cDNA encoding a human CASB81 polypeptide.
XX
SQ   Sequence 1441 BP; 337 A; 399 C; 357 G; 345 T; 3 other;


  Query Match             100.0%;  Score 72;  DB 24;  Length 1441;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 136

Qy     61 acactggttgcc 72
```

```
          ||||||||||||
Db    137 acactggttgcc 148

RESULT   7
AAA37060
ID   AAA37060 standard; cDNA; 1475 BP.
XX
AC   AAA37060;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1356 (UNQ705) cDNA sequence SEQ ID NO:133.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20111.
XX
PR   01-SEP-1998;   98US-0098716.
PR   01-SEP-1998;   98US-0098749.
PR   01-SEP-1998;   98US-0098750.
PR   02-SEP-1998;   98US-0098803.
PR   02-SEP-1998;   98US-0098821.
PR   02-SEP-1998;   98US-0098843.
PR   09-SEP-1998;   98US-0099536.
PR   09-SEP-1998;   98US-0099596.
PR   09-SEP-1998;   98US-0099598.
PR   09-SEP-1998;   98US-0099602.
PR   09-SEP-1998;   98US-0099642.
PR   10-SEP-1998;   98US-0099741.
PR   10-SEP-1998;   98US-0099754.
PR   10-SEP-1998;   98US-0099763.
PR   10-SEP-1998;   98US-0099792.
PR   10-SEP-1998;   98US-0099808.
PR   10-SEP-1998;   98US-0099812.
PR   10-SEP-1998;   98US-0099815.
PR   10-SEP-1998;   98US-0099816.
PR   15-SEP-1998;   98US-0100385.
PR   15-SEP-1998;   98US-0100388.
PR   15-SEP-1998;   98US-0100390.
PR   16-SEP-1998;   98US-0100584.
PR   16-SEP-1998;   98US-0100627.
PR   16-SEP-1998;   98US-0100661.
PR   16-SEP-1998;   98US-0100662.
PR   16-SEP-1998;   98US-0100664.
PR   17-SEP-1998;   98US-0100683.
PR   17-SEP-1998;   98US-0100684.
PR   17-SEP-1998;   98US-0100710.
PR   17-SEP-1998;   98US-0100711.
PR   17-SEP-1998;   98US-0100919.
PR   17-SEP-1998;   98US-0100930.
PR   18-SEP-1998;   98US-0100848.
PR   18-SEP-1998;   98US-0100849.
PR   18-SEP-1998;   98US-0101014.
PR   18-SEP-1998;   98US-0101068.
PR   18-SEP-1998;   98US-0101071.
PR   22-SEP-1998;   98US-0101279.
PR   23-SEP-1998;   98US-0101471.
PR   23-SEP-1998;   98US-0101472.
PR   23-SEP-1998;   98US-0101474.
PR   23-SEP-1998;   98US-0101475.
PR   23-SEP-1998;   98US-0101476.
PR   23-SEP-1998;   98US-0101477.
PR   23-SEP-1998;   98US-0101479.
PR   24-SEP-1998;   98US-0101738.
PR   24-SEP-1998;   98US-0101741.
PR   24-SEP-1998;   98US-0101743.
PR   24-SEP-1998;   98US-0101915.
PR   24-SEP-1998;   98US-0101916.
PR   29-SEP-1998;   98US-0102207.
PR   29-SEP-1998;   98US-0102240.
PR   29-SEP-1998;   98US-0102307.
PR   29-SEP-1998;   98US-0102330.
PR   29-SEP-1998;   98US-0102331.
PR   30-SEP-1998;   98US-0102484.
PR   30-SEP-1998;   98US-0102487.
PR   30-SEP-1998;   98US-0102570.
PR   30-SEP-1998;   98US-0102571.
PR   01-OCT-1998;   98US-0102684.
PR   01-OCT-1998;   98US-0102687.
PR   02-OCT-1998;   98US-0102965.
PR   06-OCT-1998;   98US-0103258.
PR   06-OCT-1998;   98US-0103449.
PR   07-OCT-1998;   98US-0103314.
PR   07-OCT-1998;   98US-0103315.
PR   07-OCT-1998;   98US-0103328.
PR   07-OCT-1998;   98US-0103395.
PR   07-OCT-1998;   98US-0103396.
PR   07-OCT-1998;   98US-0103401.
PR   08-OCT-1998;   98US-0103633.
PR   08-OCT-1998;   98US-0103678.
PR   08-OCT-1998;   98US-0103679.
PR   08-OCT-1998;   98US-0103711.
PR   14-OCT-1998;   98US-0104257.
PR   20-OCT-1998;   98US-0104987.
PR   20-OCT-1998;   98US-0105000.
PR   20-OCT-1998;   98US-0105002.
PR   21-OCT-1998;   98US-0105104.
PR   22-OCT-1998;   98US-0105169.
PR   22-OCT-1998;   98US-0105266.
PR   26-OCT-1998;   98US-0105693.
PR   26-OCT-1998;   98US-0105694.
PR   27-OCT-1998;   98US-0105807.
PR   27-OCT-1998;   98US-0105881.
PR   27-OCT-1998;   98US-0105882.
PR   27-OCT-1998;   98US-0106062.
PR   28-OCT-1998;   98US-0106023.
PR   28-OCT-1998;   98US-0106029.
PR   28-OCT-1998;   98US-0106030.
PR   28-OCT-1998;   98US-0106032.
PR   28-OCT-1998;   98US-0106033.
PR   28-OCT-1998;   98US-0106178.
PR   29-OCT-1998;   98US-0106248.
PR   29-OCT-1998;   98US-0106384.
PR   29-OCT-1998;   98US-0108500.
PR   30-OCT-1998;   98US-0106464.
PR   03-NOV-1998;   98US-0106856.
PR   03-NOV-1998;   98US-0106902.
PR   03-NOV-1998;   98US-0106905.
PR   03-NOV-1998;   98US-0106919.
PR   03-NOV-1998;   98US-0106932.
PR   03-NOV-1998;   98US-0106934.
PR   10-NOV-1998;   98US-0107783.
PR   17-NOV-1998;   98US-0108775.
PR   17-NOV-1998;   98US-0108779.
PR   17-NOV-1998;   98US-0108787.
PR   17-NOV-1998;   98US-0108788.
PR   17-NOV-1998;   98US-0108801.
PR   17-NOV-1998;   98US-0108802.
PR   17-NOV-1998;   98US-0108806.
PR   17-NOV-1998;   98US-0108807.
PR   17-NOV-1998;   98US-0108867.
PR   17-NOV-1998;   98US-0108925.
PR   18-NOV-1998;   98US-0108848.
PR   18-NOV-1998;   98US-0108849.
PR   18-NOV-1998;   98US-0108850.
PR   18-NOV-1998;   98US-0108851.
PR   18-NOV-1998;   98US-0108852.
```

```
PR   18-NOV-1998;   98US-0108858.
PR   18-NOV-1998;   98US-0108904.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker K,  Goddard A,  Gurney AL,  Smith V,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-237871/20.
DR   P-PSDB; AAY99378.
XX
PT   New mammalian DNA sequences encoding transmembrane, receptor or
PT   secreted PRO polypeptides, useful for screening of potential peptide or
PT   small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS   Claim 2; Fig 77; 773pp; English.
XX
CC   AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC   receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC   transmembrane and receptor PRO proteins can be used for screening of
CC   potential peptide or small molecule inhibitors of the relevant
CC   receptor/ligand interactions. The polypeptides and nucleotide sequences
CC   encoding then have various industrial applications, including uses as
CC   pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC   PCR primers and hybridisation probes used in the isolation of the PRO
CC   polypeptides from the present invention.
XX
SQ   Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;


  Query Match             100.0%;  Score 72;  DB 21;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy     61 acactggttgcc 72
          ||||||||||||
Db    182 acactggttgcc 193


RESULT   8
AAS15360
ID   AAS15360 standard; cDNA; 1475 BP.
XX
AC   AAS15360;
XX
DT   16-JAN-2002  (first entry)
XX
DE   cDNA encoding human PRO1356 polypeptide.
XX
KW   Human; PRO1356; clone DNA64886-1601; immune-related disorder;
KW   inflammatory disorder; infectious disorder; immunodeficiency disorder;
KW   autoimmune disorder; renal disease; demyelinating disease; skin disease;
KW   neoplasia; transplantation associated disease; gene therapy;
KW   immunosuppressive; anti-inflammatory; antiasthmatic; antidiabetic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             122..814
FT                   /*tag=  a
FT                   /product= "PRO1356 polypeptide"
FT   sig_peptide     122..193
FT                   /*tag=  b
FT   mat_peptide     194..811
FT                   /*tag=  c
XX
PN   WO200166740-A2.
XX
PD   13-SEP-2001.
XX
PF   01-MAR-2001; 2001WO-US06666.
XX
PR   03-MAR-2000; 2000US-187202P.
PR   21-MAR-2000; 2000US-191015P.
PR   30-MAY-2000; 2000WO-US14941.
PR   05-JUN-2000; 2000US-209832P.
PR   24-AUG-2000; 2000WO-US23328.
PR   01-DEC-2000; 2000WO-US32678.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Fong S,  Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL;
PI   Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-625876/72.
DR   P-PSDB; AAU09178.
XX
PT   Nucleic acids encoding PRO polypeptides, useful for detecting and
PT   treating immune related diseases and disorders in mammals including
PT   autoimmune diseases, inflammatory diseases and asthma -
XX
PS   Claim 2; Fig 1; 122pp; English.
XX
CC   The present invention relates to the isolation of 9 novel human PRO
CC   polypeptides (AAU09178-AAU09186) and the cDNA sequences encoding them.
CC   The novel PRO polypeptides include PRO1356, PRO1268, PRO1884, PRO3444,
CC   PRO3151, PRO4322, PRO9964, PRO10008 and PRO19598. The cDNA sequences
CC   encoding these PRO polypeptides have been designated as clones
CC   DNA64886-1601, DNA64903-1553, DNA84318-2520, DNA87997, DNA89273,
CC   DNA92223-2567, DNA96973, DNA101921 and DNA145887 respectively.
CC   Compositions (e.g. vaccines) containing PRO polypeptides and methods of
CC   using these compositions are useful in the treatment and diagnosis of
CC   immune-related disorders. Such disorders include immune-mediated
CC   inflammatory disorders (e.g. osteoarthritis), non-immune-mediated
CC   inflammatory  disorders (e.g. diabetes mellitus), infectious disorders
CC   (e.g. granulomatous hepatitis), immunodeficiency disorders (e.g. AIDS),
CC   autoimmune disorders (e.g. rheumatoid arthritis), immune-related renal
CC   diseases (e.g. cirrhosis), demyelinating diseases of the peripheral or
CC   central nervous system (e.g. Guillain-Barre syndrome), immune-mediated
CC   skin diseases (e.g. contact dermatitis), neoplasias and transplantation
CC   associated diseases. The polynucleotide sequences of the invention may
CC   be used in gene therapy. AAS15360-AAS15368 represent cDNA sequences
CC   encoding for the novel human PRO polypeptides of the invention.
XX
SQ   Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;


  Query Match             100.0%;  Score 72;  DB 22;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy     61 acactggttgcc 72
          ||||||||||||
Db    182 acactggttgcc 193


RESULT   9
AAS21489
ID   AAS21489 standard; cDNA; 1475 BP.
XX
AC   AAS21489;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human cDNA sequence encoding for PRO1356 polypeptide.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
```

KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200140466-A2.
XX
PD 07-JUN-2001.
XX
PF 01-DEC-2000; 2000WO-US32678.
XX
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 09-DEC-1999; 99US-0170262.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 30-DEC-1999; 99WO-US31243.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 10-NOV-2000; 2000WO-US30873.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2001-408281/43.
DR P-PSDB; AAU12417.
XX
PT Isolated, secretory and transmembrane PRO polypeptide used to detect
PT other PRO polypeptides, link bioactive molecules to cells expressing
PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT lung, breast, prostate, cervical -
XX
PS Claim 3; Fig 491; 813pp; English.
XX
CC AAS21244-AAS21518 encode for novel human secretory and transmembrane
CC PRO polypeptides. The PRO polypeptides are useful to detect other
CC PRO polypeptides, to link bioactive molecules to cells expressing
CC PRO polypeptides, to modulate biological activities of cells expressing
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample.
CC Some of the 275 sequences are also useful to stimulate the release of
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC proliferation or differentiation of chondrocytes, the proliferation or
CC gene expression in pericyte cells, the release of proteoglycans from
CC cartilage, the proliferation of inner ear utricular supporting cells or
CC of T-lymphocytes, the release of a cytokine from peripheral blood
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC to factor VIIA. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy.
XX
SQ Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;

Query Match 100.0%; Score 72; DB 22; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy 61 acactggttgcc 72
||||||||||||
Db 182 acactggttgcc 193

RESULT 10
AAF92097
ID AAF92097 standard; cDNA; 1475 BP.
XX
AC AAF92097;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1356 cDNA.
XX
KW Human; PRO protein; mapping; ss.
XX
OS Homo sapiens.
XX
PN WO200116318-A2.
XX
PD 08-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-US23328.
XX
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
PA (GETH) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI; 2001-183260/18.
DR P-PSDB; AAB87565.
XX
PT Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping. -
XX
PS Claim 2; Fig 79; 278pp; English.
XX
CC The present sequence is the coding sequence for a human PRO polypeptide
CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
CC antagonists or anti-PRO antibodies are useful for preparation of a
CC medicament useful in the treatment of a condition which is responsive to
CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
CC protein may also be employed as molecular weight markers for protein
CC electrophoresis. The PRO coding sequence has applications in molecular

```
CC   biology, including use as hybridisation probes, and in chromosome and
CC   gene mapping.
XX
SQ   Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;


  Query Match             100.0%;  Score 72;  DB 22;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy     61 acactggttgcc 72
          ||||||||||||
Db    182 acactggttgcc 193


RESULT  11
AAF54296
ID   AAF54296 standard; DNA; 1475 BP.
XX
AC   AAF54296;
XX
DT   02-APR-2001  (first entry)
XX
DE   DNA encoding protein of the invention #39.
XX
KW   Secreted; transmembrane; gene therapy; ss.
XX
OS   Unidentified.
XX
PN   WO200078961-A1.
XX
PD   28-DEC-2000.
XX
PF   18-FEB-2000; 2000WO-US04342.
XX
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   16-DEC-1999;   99WO-US30095.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Botstein D,  Desnoyers L,  Eaton DL,  Ferrara N,  Fong S;
PI   Gao W,  Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Pan J,  Paoni NF,  Roy MA,  Smith V,  Stewart TA,  Tumas D;
PI   Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2001-071395/08.
XX
PT   Secreted and transmembrane proteins and nucleic acids designated PRO,
PT   useful as hybridization probes, in chromosome and gene mapping and gene
PT   therapy -
XX
PS   Claim 2; Fig 77; 787pp; English.
XX
CC   The present invention relates to secreted and transmembrane proteins.
CC   These proteins and the DNA encoding them may be used as hybridization
CC   probes, in chromosome and gene mapping and in the generation of
CC   anti-sense RNA and DNA. They may also be used used to generate either
CC   transgenic animals or knockout animals which are in turn useful for
CC   development and screening of therapeutically useful reagents.
CC   The nucleic acids may also be used in gene therapy.
XX
SQ   Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;


  Query Match             100.0%;  Score 72;  DB 22;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy     61 acactggttgcc 72
          ||||||||||||
Db    182 acactggttgcc 193


RESULT  12
AAD08497
ID   AAD08497 standard; cDNA; 1481 BP.
XX
AC   AAD08497;
XX
DT   09-AUG-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 10 cDNA clone HTPIH83, SEQ ID NO:20.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             118..810
FT                   /*tag=  a
FT                   /product= "Human secreted protein precursor"
FT   sig_peptide     118..189
FT                   /*tag=  b
FT   mat_peptide     190..807
FT                   /*tag=  c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200136432-A2.
XX
PD   25-MAY-2001.
XX
PF   15-NOV-2000; 2000WO-US31162.
XX
PR   19-NOV-1999;   99US-0166415.
PR   30-JUN-2000; 2000US-0215136.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Baker KP,  Young PE;
XX
DR   WPI; 2001-343793/36.
DR   P-PSDB; AAE04207.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; Page 410; 509pp; English.
XX
CC   AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC   protein genes, and AAE04199-AAE04239 represent the proteins they encode.
```

```
CC   AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   18 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin ageing due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 1481 BP; 353 A; 405 C; 375 G; 348 T; 0 other;


  Query Match             100.0%;  Score 72;  DB 22;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 177

Qy      61 acactggttgcc 72
           ||||||||||||
Db     178 acactggttgcc 189


RESULT  13
AAX97818
ID   AAX97818 standard; cDNA; 1524 BP.
XX
AC   AAX97818;
XX
DT   23-SEP-1999  (first entry)
XX
DE   Human secreted protein encoding cDNA #6.
XX
KW   Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW   diagnostic; gene therapy; chromosome mapping; secretion vector; ss.
XX
OS   Homo sapiens.
XX
PN   WO9925825-A2.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-IB01862.
XX
PR   04-SEP-1998;   98US-0099273.
PR   13-NOV-1997;   97US-0066677.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-347472/29.
DR   P-PSDB; AAY36134.
XX
PT   Extended cDNAs encoding secreted proteins
XX
PS   Example 28; Page 155-156; 307pp; English.
XX
CC   AAX97813-X97906 represent extended cDNA's which encode novel human
CC   secreted proteins (see AAY36129-Y36222) and which have cytostatic,
CC   thrombotic and osteopathic activity. The extended cDNAs can be used to
CC   express secreted proteins or parts of them or to obtain antibodies
CC   capable of binding to the secreted proteins. They may also be used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   Uses also include design of expression vectors and secretion vectors.
XX
SQ   Sequence 1524 BP; 365 A; 404 C; 382 G; 353 T; 20 other;


  Query Match             100.0%;  Score 72;  DB 20;  Length 1524;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-15;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 219

Qy      61 acactggttgcc 72
           ||||||||||||
Db     220 acactggttgcc 231


RESULT  14
AAI58013
ID   AAI58013 standard; cDNA; 1540 BP.
XX
AC   AAI58013;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 216.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
```

XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM38857.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 216; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1540 BP; 365 A; 419 C; 389 G; 367 T; 0 other;

Query Match 100.0%; Score 72; DB 22; Length 1540;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 183 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 242

Qy 61 acactggttgcc 72
||||||||||||
Db 243 acactggttgcc 254

RESULT 15
AAI59799/c
ID AAI59799 standard; cDNA; 1761 BP.
XX
AC AAI59799;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 3788.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM40643.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 3788; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1761 BP; 407 A; 446 C; 478 G; 430 T; 0 other;

Query Match 100.0%; Score 72; DB 22; Length 1761;
Best Local Similarity 100.0%; Pred. No. 3e-15;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1641 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 1582

Qy 61 acactggttgcc 72
||||||||||||
Db 1581 ACACTGGTTGCC 1570

Search completed: June 2, 2002, 20:09:03
Job time: 14748 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:19:26 ; Search time 118.89 Seconds
(without alignments)
214.879 Million cell updates/sec

Title: US-09-663-600A-186
Perfect score: 1174
Sequence: 1 MASLGLQLVGYILGLLGLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1174	100.0	230	20	AAY36181	Human secreted pro
2	1174	100.0	230	21	AAY99378	Human PRO1356 (UNQ
3	1174	100.0	230	21	AAY84609	A human membrane a
4	1174	100.0	230	22	AAU09178	Human PRO1356 poly
5	1174	100.0	230	22	AAU12417	Human PRO1356 poly
6	1174	100.0	230	22	AAE04207	Human gene 10 enco
7	1174	100.0	230	22	AAB87565	Human PRO1356. Ho
8	1174	100.0	230	22	AAB88342	Human membrane or
9	1174	100.0	230	22	AAB66127	Protein of the inv
10	1169	99.6	230	22	AAM38857	Human polypeptide
11	1169	99.6	260	22	AAM40643	Human polypeptide
12	1156	98.5	230	20	AAY36134	Human secreted pro
13	1096	93.4	230	21	AAY51676	Murine clodin 2 pr
14	695	59.2	140	22	AAE04228	Human gene 10 enco
15	635	54.1	155	21	AAB54052	Human pancreatic c
16	571	48.6	239	21	AAY99431	Human PRO1571 (UNQ
17	571	48.6	239	22	AAU29200	Human PRO polypept
18	571	48.6	239	22	AAB66180	Protein of the inv
19	570	48.6	113	22	AAE04281	Human gene 10 enco
20	559	47.6	114	20	AAY12226	Human 5' EST secre
21	529	45.1	233	22	AAM78559	Human protein SEQ
22	515	43.9	233	22	AAM79543	Human protein SEQ
23	455	38.8	210	20	AAW86307	Kidney injury asso
24	450	38.3	220	20	AAY13939	Human transmembran
25	450	38.3	220	22	ABB50279	Claudin 3 ovarian
26	448	38.2	88	22	AAG75581	Human colon cancer
27	446	38.0	208	21	AAB52100	Gene 48 human secr
28	443	37.7	219	21	AAY51681	Murine clodin 6 pr
29	442.5	37.7	209	21	AAB43133	Human ORFX ORF2897
30	442.5	37.7	209	22	ABB50278	Claudin 4 ovarian
31	442.5	37.7	210	22	AAG75467	Human colon cancer
32	441	37.6	217	22	AAB64401	Amino acid sequenc
33	441	37.6	219	21	AAY51679	Murine clodin 3 pr
34	438	37.3	220	21	AAY99434	Human PRO1488 (UNQ
35	438	37.3	220	22	AAB66183	Protein of the inv
36	438	37.3	220	22	AAB50968	Human PRO1488 prot
37	437	37.2	220	22	AAM39937	Human polypeptide
38	437	37.2	220	22	AAB88319	Human membrane or
39	437	37.2	235	22	AAM41723	Human polypeptide
40	434.5	37.0	254	22	ABG27980	Novel human diagno
41	431.5	36.8	211	21	AAY51682	Murine clodin 7 pr
42	430	36.6	86	22	AAE04229	Human gene 10 enco
43	427.5	36.4	205	21	AAB52101	Human secreted pro
44	427.5	36.4	210	19	AAW63697	Human secreted pro
45	427.5	36.4	211	21	AAY57609	Human apoptosis as

ALIGNMENTS

RESULT 1
AAY36181
ID AAY36181 standard; protein; 230 AA.
XX
AC AAY36181;
XX
DT 23-SEP-1999 (first entry)
XX
DE Human secreted protein #53.
XX
KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW diagnostic; gene therapy; chromosome mapping; secretion vector.
XX
OS Homo sapiens.
XX
PN WO9925825-A2.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-IB01862.
XX
PR 04-SEP-1998; 98US-0099273.
PR 13-NOV-1997; 97US-0066677.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
PR 10-AUG-1998; 98US-0096116.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-347472/29.
DR N-PSDB; AAX97865.

XX
PT Extended cDNAs encoding secreted proteins
XX
PS Claim 7; Page 289; 307pp; English.
XX
CC AAY36129-Y36222 represent novel human secreted proteins encoded by the
CC extended cDNA sequences represented in AAX97813-X97906. The proteins
CC of the invention have cytostatic, thrombotic and osteopathic activity.
CC The extended cDNAs can be used to express secreted proteins or parts of
CC them or to obtain antibodies capable of binding to the secreted
CC proteins. They may also be used in diagnostic, forensic, gene therapy
CC and chromosome mapping procedures. Uses also include design of
CC expression vectors and secretion vectors.
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 1174; DB 20; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.1e-117;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 maslglqlvgyilgllglltlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230
```

RESULT 2
AAY99378
ID AAY99378 standard; Protein; 230 AA.
XX
AC AAY99378;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1356 (UNQ705) amino acid sequence SEQ ID NO:134.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.

```
PR   27-OCT-1998;   98US-0106062.
PR   28-OCT-1998;   98US-0106023.
PR   28-OCT-1998;   98US-0106029.
PR   28-OCT-1998;   98US-0106030.
PR   28-OCT-1998;   98US-0106032.
PR   28-OCT-1998;   98US-0106033.
PR   28-OCT-1998;   98US-0106178.
PR   29-OCT-1998;   98US-0106248.
PR   29-OCT-1998;   98US-0106384.
PR   29-OCT-1998;   98US-0108500.
PR   30-OCT-1998;   98US-0106464.
PR   03-NOV-1998;   98US-0106856.
PR   03-NOV-1998;   98US-0106902.
PR   03-NOV-1998;   98US-0106905.
PR   03-NOV-1998;   98US-0106919.
PR   03-NOV-1998;   98US-0106932.
PR   03-NOV-1998;   98US-0106934.
PR   10-NOV-1998;   98US-0107783.
PR   17-NOV-1998;   98US-0108775.
PR   17-NOV-1998;   98US-0108779.
PR   17-NOV-1998;   98US-0108787.
PR   17-NOV-1998;   98US-0108788.
PR   17-NOV-1998;   98US-0108801.
PR   17-NOV-1998;   98US-0108802.
PR   17-NOV-1998;   98US-0108806.
PR   17-NOV-1998;   98US-0108807.
PR   17-NOV-1998;   98US-0108867.
PR   17-NOV-1998;.  98US-0108925.
PR   18-NOV-1998;   98US-0108848.
PR   18-NOV-1998;   98US-0108849.
PR   18-NOV-1998;   98US-0108850.
PR   18-NOV-1998;   98US-0108851.
PR   18-NOV-1998;   98US-0108852.
PR   18-NOV-1998;   98US-0108858.
PR   18-NOV-1998;   98US-0108904.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker K,  Goddard A,  Gurney AL,  Smith V,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-237871/20.
DR   N-PSDB; AAA37060.
XX
PT   New mammalian DNA sequences encoding transmembrane, receptor or
PT   secreted PRO polypeptides, useful for screening of potential peptide or
PT   small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS   Claim 12; Fig 78; 773pp; English.
XX
CC   AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC   receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC   transmembrane and receptor PRO proteins can be used for screening of
CC   potential peptide or small molecule inhibitors of the relevant
CC   receptor/ligand interactions. The polypeptides and nucleotide sequences
CC   encoding then have various industrial applications, including uses as
CC   pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC   PCR primers and hybridisation probes used in the isolation of the PRO
CC   polypeptides from the present invention.
XX
SQ   Sequence   230 AA;


  Query Match             100.0%;  Score 1174;  DB 21;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-117;
  Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230


RESULT   3
AAY84609
ID   AAY84609 standard; Protein; 230 AA.
XX
AC   AAY84609;
XX
DT   25-JUL-2000  (first entry)
XX
DE   A human membrane associated organizational protein (HJNCT).
XX
KW   Human; membrane associated organizational protein; HJNCT;
KW   cell proliferative disorder; cancer; autoimmune disorder;
KW   inflammatory disorder; neurological disorder; developmental disorder;
KW   vesicle trafficking; reproductive disorder; gastrointestinal disorder;
KW   renal disorder; atherosclerosis; leukaemia; rheumatoid arthritis;
KW   Alzheimer's disease; anxiety; diabetes; ovulatory defect; renal failure;
KW-  irritable bowel syndrome; allergy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   29
FT                   /note= "potential phosphorylation site"
FT   Modified-site   62
FT                   /note= "potential phosphorylation site"
FT   Domain          117..138
FT                   /note= "transmembrane domain"
FT   Modified-site   155
FT                   /note= "potential phosphorylation site"
FT   Domain          164..182
FT                   /note= "transmembrane domain"
FT   Modified-site   187
FT                   /note= "potential phosphorylation site"
FT   Modified-site   190
FT                   /note= "potential glycosylation site"
FT   Modified-site   208
FT                   /note= "potential phosphorylation site"
FT   Modified-site   224
FT                   /note= "potential phosphorylation site"
XX
PN   WO200018915-A2.
XX
PD   06-APR-2000.
XX
PF   23-SEP-1999;   99WO-US22082.
XX
PR   25-SEP-1998;   98US-0155215.
PR   13-OCT-1998;   98US-0155251.
PR   04-MAY-1999;   99US-0172228.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Yue H,  Lal P,  Corley NC,  Guegler KJ,  Baughn MR,  Lu AD,  Tang YT;
XX
DR   WPI; 2000-293154/25.
DR   N-PSDB; AAA12585.
XX
PT   Human membrane associated organizational protein and nucleic acid
PT   sequences useful in the diagnosis, treatment and prevention of cell
PT   proliferative associated disorders e.g. cancer, rheumatoid arthritis
PT   and Alzheimer's disease -
XX
PS   Claim 1; Page 76-77; 84pp; English.
```

```
XX
CC   The present sequence represents a membrane associated organizational
CC   protein (HJNCT). HJNCT is used for the diagnosis, treatment and
CC   prevention of cell proliferative disorders including cancer and
CC   autoimmune/inflammatory, neurological, developmental, vesicle
CC   trafficking, reproductive, gastrointestinal and renal disorders. These
CC   disorders may include atherosclerosis, leukaemia, allergies, rheumatoid
CC   arthritis, Alzheimer's disease, anxiety, diabetes, ovulatory defects,
CC   renal failure and irritable bowel syndrome. A vector expressing HJNCT,
CC   and an agonist of HJNCT can be used to treat or prevent a disorder
CC   associated with decreased expression or activity of HJNCT. An antagonist
CC   of HJNCT or a vector expressing the complement of a polynucleotide
CC   encoding HJNCT can be used to treat or prevent a disorder associated
CC   with increased expression or activity of HJNCT. Antibodies which bind
CC   HJNCT can be used for diagnosis of disorders associated with HJNCT
CC   expression or to monitor patients being treated with HJNCT, agonists,
CC   antagonists or inhibitors of HJNCT. Assays are preferably carried out on
CC   body fluids from a patient using radioimmunoassay, enzyme linked
CC   immunosorbent assays or fluorescent activated cell sorting assays.
CC   Polynucleotides encoding HJNCT are also used in hybridisation assays to
CC   determine absence, presence or excess expression of HJNCT and to monitor
CC   regulation of HJNCT levels during disease therapy.
XX
SQ   Sequence   230 AA;

  Query Match             100.0%;  Score 1174;  DB 21;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-117;
  Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230


RESULT   4
AAU09178
ID   AAU09178 standard; Protein; 230 AA.
XX
AC   AAU09178;
XX
DT   16-JAN-2002  (first entry)
XX
DE   Human PRO1356 polypeptide.
XX
KW   Human; PRO1356; clone DNA64886-1601; immune-related disorder;
KW   inflammatory disorder; infectious disorder; immunodeficiency disorder;
KW   autoimmune disorder; renal disease; demyelinating disease; skin disease;
KW   neoplasia; transplantation associated disease; immunosuppressive;
KW   anti-inflammatory; antiasthmatic; antidiabetic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..24
FT                   /label= Signal_peptide
FT   Modified-site   17..23
FT                   /note= "N-myristoylation site"
FT   Modified-site   20..26
FT                   /note= "N-myristoylation site"
FT   Protein         25..230
FT                   /label= Mature_PRO1356_polypeptide
FT   Region          46..60
FT                   /note= "Also found in PMP-22, EMP and MP20 family
FT                   proteins"
FT   Modified-site   60..66
FT                   /note= "N-myristoylation site"
FT   Domain          82..102
FT                   /label= Transmembrane_domain
FT   Modified-site   101..107
FT                   /note= "N-myristoylation site"
FT   Domain          117..140
FT                   /label= Transmembrane_domain
FT   Modified-site   128..134
FT                   /note= "N-myristoylation site"
FT   Domain          163..182
FT                   /label= Transmembrane_domain
FT   Modified-site   167..173
FT                   /note= "N-myristoylation site"
FT   Modified-site   178..184
FT                   /note= "N-myristoylation site"
FT   Modified-site   190..194
FT                   /note= "N-glycosylation site"
FT   Modified-site   216..225
FT                   /note= "Tyrosine kinase phosphorylation site"
XX
PN   WO200166740-A2.
XX
PD   13-SEP-2001.
XX
PF   01-MAR-2001; 2001WO-US06666.
XX
PR   03-MAR-2000; 2000US-187202P.
PR   21-MAR-2000; 2000US-191015P.
PR   30-MAY-2000; 2000WO-US14941.
PR   05-JUN-2000; 2000US-209832P.
PR   24-AUG-2000; 2000WO-US23328.
PR   01-DEC-2000; 2000WO-US32678.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Fong S,  Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL;
PI   Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-625876/72.
DR   N-PSDB; AAS15360.
XX
PT   Nucleic acids encoding PRO polypeptides, useful for detecting and
PT   treating immune related diseases and disorders in mammals including
PT   autoimmune diseases, inflammatory diseases and asthma -
XX
PS   Claim 10; Fig 2; 122pp; English.
XX
CC   The present invention relates to the isolation of 9 novel human PRO
CC   polypeptides and the cDNA sequences (AAS15360-AAS15368) encoding them.
CC   The novel PRO polypeptides include PRO1356, PRO1268, PRO1884, PRO3444,
CC   PRO3151, PRO4322, PRO9964, PRO10008 and PRO19598. The cDNA sequences
CC   encoding these PRO polypeptides have been designated as clones
CC   DNA64886-1601, DNA64903-1553, DNA84318-2520, DNA87997, DNA89273,
CC   DNA92223-2567, DNA96973, DNA101921 and DNA145887 respectively.
CC   Compositions (e.g. vaccines) containing PRO polypeptides and methods of
CC   using these compositions are useful in the treatment and diagnosis of
CC   immune-related disorders. Such disorders include immune-mediated
CC   inflammatory disorders (e.g. osteoarthritis), non-immune-mediated
CC   inflammatory  disorders (e.g. diabetes mellitus), infectious disorders
CC   (e.g. granulomatous hepatitis), immunodeficiency disorders (e.g. AIDS),
CC   autoimmune disorders (e.g. rheumatoid arthritis), immune-related renal
CC   diseases (e.g. cirrhosis), demyelinating diseases of the peripheral or
CC   central nervous system (e.g. Guillain-Barre syndrome), immune-mediated
CC   skin diseases (e.g. contact dermatitis), neoplasias and transplantation
CC   associated diseases. The polynucleotide sequences of the invention may
CC   be used in gene therapy. AAU09178-AAU09186 represent the novel human
CC   PRO polypeptides of the invention.
XX
```

```
SQ   Sequence   230 AA;

  Query Match             100.0%;  Score 1174;  DB 22;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-117;
  Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy    121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy    181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230


RESULT   5
AAU12417
ID   AAU12417 standard; Protein; 230 AA.
XX
AC   AAU12417;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human PRO1356 polypeptide sequence.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   09-DEC-1999;   99US-0170262.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Beresini M,  Deforge L,  Desnoyers L,  Filvaroff E,  Gao W;
PI   Gerritsen ME,  Goddard A,  Godowski PJ,  Gurney AL,  Sherwood S;
PI   Smith V,  Stewart TA,  Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-408281/43.
DR   N-PSDB; AAS21489.
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical -
XX
PS   Claim 12; Fig 492; 813pp; English.
XX
CC   AAU12172-AAU12446 represent novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence   230 AA;

  Query Match             100.0%;  Score 1174;  DB 22;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-117;
  Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy    121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy    181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230


RESULT   6
AAE04207
ID   AAE04207 standard; Protein; 230 AA.
XX
AC   AAE04207;
XX
DT   09-AUG-2001  (first entry)
XX
DE   Human gene 10 encoded secreted protein HTPIH83, SEQ ID NO:62.
XX
```

```
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive;
KW   binding partner identification.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..24
FT                   /label= signal_peptide
FT   Protein         25..230
FT                   /note= "Mature secreted protein"
XX
PN   WO200136432-A2.
XX
PD   25-MAY-2001.
XX
PF   15-NOV-2000; 2000WO-US31162.
XX
PR   19-NOV-1999;   99US-0166415.
PR   30-JUN-2000; 2000US-0215136.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Baker KP,  Young PE;
XX
DR   WPI; 2001-343793/36.
DR   N-PSDB; AAD08497.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 11; Page 439; 509pp; English.
XX
CC   AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC   protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC   AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   18 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin ageing due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein of the invention.
XX
SQ   Sequence   230 AA;
```

```
  Query Match             100.0%;  Score 1174;  DB 22;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-117;
  Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230
```

```
RESULT   7
AAB87565
ID   AAB87565 standard; Protein; 230 AA.
XX
AC   AAB87565;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human PRO1356.
XX
KW   Human; PRO protein; mapping.
XX
OS   Homo sapiens.
XX
PN   WO200116318-A2.
XX
PD   08-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-US23328.
XX
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   07-DEC-1999;   99US-0169495.
PR   09-DEC-1999;   99US-0170262.
PR   11-JAN-2000; 2000US-0175481.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   01-MAR-2000; 2000WO-US05601.
PR   03-MAR-2000; 2000US-0187202.
PR   25-APR-2000; 2000US-0199397.
PR   22-MAY-2000; 2000WO-US14042.
PR   05-JUN-2000; 2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   N-PSDB; AAF92097.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 12; Fig 80; 278pp; English.
XX
CC   The present sequence is a human PRO polypeptide (secreted and
CC   transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC   anti-PRO antibodies are useful for preparation of a medicament useful in
```

CC the treatment of a condition which is responsive to the PRO protein,
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC employed as molecular weight markers for protein electrophoresis. The PRO
CC coding sequence has applications in molecular biology, including use as
CC hybridisation probes, and in chromosome and gene mapping.
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 1174; DB 22; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.1e-117;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy    121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy    181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230
```

RESULT 8
AAB88342
ID AAB88342 standard; Protein; 230 AA.
XX
AC AAB88342;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human membrane or secretory protein clone PSEC0059.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI; 2001-093989/11.
DR N-PSDB; AAF93769.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 1; SEQ ID 52; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 1174; DB 22; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.1e-117;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy    121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy    181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230
```

RESULT 9
AAB66127
ID AAB66127 standard; protein; 230 AA.
XX
AC AAB66127;
XX
DT 02-APR-2001 (first entry)
XX
DE Protein of the invention #39.
XX
KW Secreted; transmembrane; gene therapy.
XX
OS Unidentified.
XX
PN WO200078961-A1.
XX
PD 28-DEC-2000.
XX
PF 18-FEB-2000; 2000WO-US04342.
XX
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.

```
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Botstein D,  Desnoyers L,  Eaton DL,  Ferrara N,  Fong S;
PI   Gao W,  Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Pan J,  Paoni NF,  Roy MA,  Smith V,  Stewart TA,  Tumas D;
PI   Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2001-071395/08.
XX
PT   Secreted and transmembrane proteins and nucleic acids designated PRO,
PT   useful as hybridization probes, in chromosome and gene mapping and gene
PT   therapy -
XX
PS   Claim 1; Fig 78; 787pp; English.
XX
CC   The present invention relates to secreted and transmembrane proteins.
CC   These proteins and the DNA encoding them may be used as hybridization
CC   probes, in chromosome and gene mapping and in the generation of
CC   anti-sense RNA and DNA. They may also be used used to generate either
CC   transgenic animals or knockout animals which are in turn useful for
CC   development and screening of therapeutically useful reagents.
CC   The nucleic acids may also be used in gene therapy.
XX
SQ   Sequence   230 AA;


  Query Match             100.0%;  Score 1174;  DB 22;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-117;
  Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230


RESULT  10
AAM38857
ID   AAM38857 standard; Protein; 230 AA.
XX
AC   AAM38857;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2002.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI58013.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 3; SEQ ID NO 2002; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   230 AA;


  Query Match             99.6%;  Score 1169;  DB 22;  Length 230;
  Best Local Similarity   99.6%;  Pred. No. 1.4e-116;
  Matches  229;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           |||||| ||||||||||||||||||||||||||||||||||||||||||
Db     181 lcfscscqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230


RESULT  11
AAM40643
ID   AAM40643 standard; Protein; 260 AA.
XX
AC   AAM40643;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5574.
```

XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI59799.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5574; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 260 AA;

Query Match 99.6%; Score 1169; DB 22; Length 260;
Best Local Similarity 99.6%; Pred. No. 1.7e-116;
Matches 229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
Db 31 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 90

Qy 61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
Db 91 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 150

Qy 121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
Db 151 ggvffilggllgfipvawnlhgilrdfysplvpdsmkfeigealylgiisslfsliagii 210

Qy 181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
Db 211 lcfscscqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 260

RESULT 12
AAY36134
ID AAY36134 standard; protein; 230 AA.
XX
AC AAY36134;
XX
DT 23-SEP-1999 (first entry)
XX
DE Human secreted protein #6.
XX
KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW diagnostic; gene therapy; chromosome mapping; secretion vector.
XX
OS Homo sapiens.
XX
PN WO9925825-A2.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-IB01862.
XX
PR 04-SEP-1998; 98US-0099273.
PR 13-NOV-1997; 97US-0066677.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
PR 10-AUG-1998; 98US-0096116.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-347472/29.
DR N-PSDB; AAX97818.
XX
PT Extended cDNAs encoding secreted proteins
XX
PS Example 28; Page 234-235; 307pp; English.
XX
CC AAY36129-Y36222 represent novel human secreted proteins encoded by the
CC extended cDNA sequences represented in AAX97813-X97906. The proteins
CC of the invention have cytostatic, thrombotic and osteopathic activity.
CC The extended cDNAs can be used to express secreted proteins or parts of
CC them or to obtain antibodies capable of binding to the secreted
CC proteins. They may also be used in diagnostic, forensic, gene therapy
CC and chromosome mapping procedures. Uses also include design of
CC expression vectors and secretion vectors.
XX
SQ Sequence 230 AA;

Query Match 98.5%; Score 1156; DB 20; Length 230;
Best Local Similarity 98.7%; Pred. No. 3.4e-115;
Matches 227; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
Db 1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy 61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
Db 61 itqcdiystllglpadixaaqammvtssaisslaciisvvgmxctvfcqesrakdrvava 120

Qy 121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
Db 121 ggvffilggllgfipvawnlhgilrdfysplvpdsmxfeigealylgiisslfsliagii 180

```
Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lcfscssqrnrsnyydayqaqplatrssprpgqppkvksefnsysltgyv 230


RESULT  13
AAY51676
ID   AAY51676 standard; Protein; 230 AA.
XX
AC   AAY51676;
XX
DT   02-JUN-2000  (first entry)
XX
DE   Murine clodin 2 protein.
XX
KW   Clodin 2; murine; tight junction-constituting membrane protein;
KW   medicine.
XX
OS   Mus sp.
XX
PN   JP2000032984-A.
XX
PD   02-FEB-2000.
XX
PF   26-JUN-1998;   98JP-0179847.
XX
PR   15-MAY-1998;   98JP-0133215.
XX
PA   (EISA ) EISAI CO LTD.
XX
DR   WPI; 2000-285512/25.
DR   N-PSDB; AAZ89137.
XX
PT   Tight junction-constituting membrane protein clodin family - useful in
PT   the medical field
XX
PS   Claim 2; Page 10; 22pp; Japanese.
XX
CC   This invention describes novel murine nucleic acid sequences encoding the
CC   clodin family of tight junction (TJ)-constituting membrane protein. The
CC   membrane protein can be used in medical field. This sequence represents
CC   the clodin 2 protein described in the method of the invention.
XX
SQ   Sequence   230 AA;


  Query Match             93.4%;  Score 1096;  DB 21;  Length 230;
  Best Local Similarity   91.3%;  Pred. No. 8.8e-109;
  Matches  210;  Conservative   11;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           |||||:|||||||||||||| :|||||:|:|||||||||||||||||||||||||||||
Db       1 maslgvqlvgyilgllgllgtsiamllpnwrtssyvgasivtavgfskglwmecathstg 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||:||||||||||||||:|||||||||
Db      61 itqcdiystllglpadiqaaqammvtssamsslaciisvvgmrctvfcqdsrakdrvavv 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           |||||||||:|||||||||||||||||||||||||||||||||||||:||||:||:|
Db     121 ggvffilggilgfipvawnlhgilrdfysplvpdsmkfeigealylgiisalfslvagvi 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           |||||| | ||:|||| ||||||||||||   | || ||||||||||||
Db     181 lcfscspqgnrtnyydgyqaqplatrssprsaqqpkaksefnsysltgyv 230


RESULT  14
AAE04228
ID   AAE04228 standard; Protein; 140 AA.
XX
AC   AAE04228;
XX
DT   09-AUG-2001  (first entry)
XX
DE   Human gene 10 encoded secreted protein HTPIH83, SEQ ID NO:83.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive;
KW   binding partner identification.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..24
FT                   /label= signal_peptide
FT   Protein         25..140
FT                   /note= "Mature secreted protein"
FT   Misc-difference 136
FT                   /note= "Encoded by GNT"
XX
PN   WO200136432-A2.
XX
PD   25-MAY-2001.
XX
PF   15-NOV-2000; 2000WO-US31162.
XX
PR   19-NOV-1999;   99US-0166415.
PR   30-JUN-2000; 2000US-0215136.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Baker KP,  Young PE;
XX
DR   WPI; 2001-343793/36.
DR   N-PSDB; AAD08518.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 11; Page 459; 509pp; English.
XX
CC   AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC   protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC   AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   18 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin ageing due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
```

```
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein of the invention.
XX
SQ   Sequence   140 AA;


  Query Match             59.2%;  Score 695;  DB 22;  Length 140;
  Best Local Similarity   98.6%;  Pred. No. 3.2e-66;
  Matches  138;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 itqcdiystllglpadiqaaqammvtssaisslaciisvvgmrctvfcqesrakdrvava 120

Qy     121 GGVFFILGGLLGFIPVAWNL 140
           |||||||| |||||| ||||
Db     121 ggvffilgsllgfipxawnl 140


RESULT  15
AAB54052
ID   AAB54052 standard; Protein; 155 AA.
XX
AC   AAB54052;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen protein sequence SEQ ID NO:504.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   N-PSDB; AAC98817.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 11; Page 942-943; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   155 AA;


  Query Match             54.1%;  Score 635;  DB 21;  Length 155;
  Best Local Similarity   97.7%;  Pred. No. 9.4e-60;
  Matches  127;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 maslglqlvgyilgllgllgtlvamllpswktssyvgasivtavgfskglwmecathstg 85

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           |||||||||||||||||| |||||||||||| |||||||||||||||||||||||||||
Db      86 itqcdiystllglpadiqxaqammvtssaisxlaciisvvgmrctvfcqesrakdrvava 145

Qy     121 GGVFFILGGL 130
           |||||||| |
Db     146 ggvffilgxl 155


Search completed: June  2, 2002, 20:10:58
Job time: 6692 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 20:00:07 ; Search time 3664.46 Seconds
(without alignments)
265.191 Million cell updates/sec

Title: US-09-663-600A-139_COPY_36_107
Perfect score: 72
Sequence: 1 atggcctctcttggcctcca.....ttttgggcacactggttgcc 72

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	72	100.0	369	10	T86266	T86266 yd84h07.r1
2	72	100.0	494	10	BG385562	BG385562 602453858
3	72	100.0	643	10	BG328625	BG328625 602427889
4	72	100.0	822	10	BG386074	BG386074 602455248
5	72	100.0	945	10	BE513091	BE513091 601171545
6	70.4	97.8	808	10	BG469035	BG469035 602510779
7	60.8	84.4	450	9	AV602537	AV602537 AV602537
8	60.8	84.4	517	10	BE480970	BE480970 166244 BA
9	60.8	84.4	551	9	AV605969	AV605969 AV605969
10	60.8	84.4	555	9	AV604013	AV604013 AV604013
11	60.8	84.4	569	10	BE667438	BE667438 151799 MA
12	60	83.3	728	10	BG325755	BG325755 602424466
13	60	83.3	856	10	BF541534	BF541534 602067906
14	60	83.3	886	10	BF032123	BF032123 601559727
15	60	83.3	978	10	BG164062	BG164062 602341087
16	56.6	78.6	601	9	AV647382	AV647382 AV647382
17	56	77.8	348	9	BB843637	BB843637 BB843637
18	56	77.8	371	9	BB873329	BB873329 BB873329
19	56	77.8	383	9	BB843776	BB843776 BB843776
20	56	77.8	403	9	AW259767	AW259767 um77h09.y
21	56	77.8	422	9	BB848306	BB848306 BB848306
22	56	77.8	433	9	BB845010	BB845010 BB845010
23	56	77.8	483	9	AI116814	AI116814 ue21c11.y
24	56	77.8	490	9	AW107793	AW107793 ul93h07.y
25	56	77.8	493	9	AI115927	AI115927 ue25f09.y
26	56	77.8	501	9	AW258786	AW258786 um73d04.y
27	56	77.8	520	9	AI746647	AI746647 ul01g04.y
28	56	77.8	529	9	AI119108	AI119108 ue23d11.y
29	56	77.8	593	9	AI789119	AI789119 uk51c01.y
30	56	77.8	619	9	BB653578	BB653578 BB653578
31	56	77.8	628	9	AI788398	AI788398 uk58c09.y
32	56	77.8	659	9	AI789490	AI789490 uk99h03.y
33	56	77.8	673	9	AI746723	AI746723 ul06f09.y
34	56	77.8	722	9	AI790813	AI790813 uk28c10.y
35	56	77.8	750	9	AW475316	AW475316 un64d02.y
36	56	77.8	2829	11	AK004990	AK004990 Mus muscu
37	54.4	75.6	279	9	BB574507	BB574507 BB574507
38	54.4	75.6	458	10	BF789570	BF789570 602103665
39	54.4	75.6	704	10	BI330754	BI330754 602982419
40	54.4	75.6	727	10	BI101652	BI101652 602887470
41	54.4	75.6	757	10	BI143586	BI143586 602907418
42	54.4	75.6	758	10	BI102679	BI102679 602888236
43	54.4	75.6	777	10	BI102100	BI102100 602887847
44	54.4	75.6	846	10	BF789255	BF789255 602105156
45	54.4	75.6	1114	10	BF789836	BF789836 602105395

ALIGNMENTS

RESULT 1
T86266
LOCUS T86266 369 bp mRNA linear EST 17-MAR-1995
DEFINITION yd84h07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:114973 5' similar to SP:A39484 A39484 ANDROGEN-WITHDRAWAL APOPTOSIS PROTEIN RVP1, PROSTATIC - ;, mRNA sequence.
ACCESSION T86266
VERSION T86266.1 GI:714618
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 369)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1623
High quality sequence stops: 288 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1623 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 288.
FEATURES Location/Qualifiers
source 1. .369
/organism="Homo sapiens"
/db_xref="GDB:470590"
/db_xref="taxon:9606"
/clone="IMAGE:114973"
/clone_lib="Soares fetal liver spleen 1NFLS"

/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 73 a 106 c 100 g 87 t 3 others
ORIGIN.

Query Match 100.0%; Score 72; DB 10; Length 369;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 77 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 136

Qy 61 acactggttgcc 72
||||||||||||
Db 137 ACACTGGTTGCC 148

RESULT 2
BG385562
LOCUS BG385562 494 bp mRNA linear EST 12-MAR-2001
DEFINITION 602453858F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582214 5', mRNA sequence.
ACCESSION BG385562
VERSION BG385562.1 GI:13278350
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1304 row: c column: 15
High quality sequence stop: 493.
FEATURES Location/Qualifiers
source 1. .494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4582214"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 91 a 145 c 132 g 126 t
ORIGIN

Query Match 100.0%; Score 72; DB 10; Length 494;
Best Local Similarity 100.0%; Pred. No. 4.6e-12;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 94

Qy 61 acactggttgcc 72
||||||||||||
Db 95 ACACTGGTTGCC 106

RESULT 3
BG328625
LOCUS BG328625 643 bp mRNA linear EST 27-FEB-2001
DEFINITION 602427889F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4547230 5', mRNA sequence.
ACCESSION BG328625
VERSION BG328625.1 GI:13135063
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 643)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1235 row: a column: 23
High quality sequence stop: 641.
FEATURES Location/Qualifiers
source 1. .643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4547230"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 130 a 173 c 177 g 163 t
ORIGIN

Query Match 100.0%; Score 72; DB 10; Length 643;
Best Local Similarity 100.0%; Pred. No. 4.9e-12;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 116 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 175

Qy 61 acactggttgcc 72
||||||||||||

```
Db    176 ACACTGGTTGCC 187


RESULT   4
BG386074
LOCUS       BG386074                822 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602455248F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583675 5',
            mRNA sequence.
ACCESSION   BG386074
VERSION     BG386074.1  GI:13279520
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 822)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1307  row: p  column: 12
            High quality sequence stop: 746.
FEATURES             Location/Qualifiers
     source          1. .822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4583675"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      181 a    224 c    217 g    200 t
ORIGIN


  Query Match             100.0%;  Score 72;  DB 10;  Length 822;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-12;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 186

Qy     61 acactggttgcc 72
          ||||||||||||
Db    187 ACACTGGTTGCC 198


RESULT   5
BE513091
LOCUS       BE513091                945 bp    mRNA    linear   EST 07-AUG-2000
DEFINITION  601171545F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:3545171 5',
            mRNA sequence.
ACCESSION   BE513091
VERSION     BE513091.1  GI:9720302
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM241  row: i  column: 12
            High quality sequence start: 36
            High quality sequence stop: 782.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3545171"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      194 a    262 c    263 g    226 t
ORIGIN


  Query Match             100.0%;  Score 72;  DB 10;  Length 945;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-12;
  Matches   72;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 246

Qy     61 acactggttgcc 72
          ||||||||||||
Db    247 ACACTGGTTGCC 258


RESULT   6
BG469035
LOCUS       BG469035                808 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602510779F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4645075 5',
            mRNA sequence.
ACCESSION   BG469035
VERSION     BG469035.1  GI:13401220
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
```

```
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1417  row: n  column: 20
            High quality sequence stop: 585.
FEATURES              Location/Qualifiers
     source           1. .808
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4645075"
                      /clone_lib="NIH_MGC_15"
                      /tissue_type="adenocarcinoma cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Size-selected >500bp for average
                      insert size 1.8kb.  Library constructed by Ling Hong in
                      the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using  ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      204 a    216 c    200 g    188 t
ORIGIN


  Query Match             97.8%;  Score 70.4;  DB 10;  Length 808;
  Best Local Similarity   98.6%;  Pred. No. 1.6e-11;
  Matches   71;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 ATGGACTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 183

Qy     61 acactggttgcc 72
          ||||||||||||
Db    184 ACACTGGTTGCC 195


RESULT   7
AV602537
LOCUS       AV602537                450 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  AV602537 Bos taurus kidney fetus Bos taurus cDNA clone E1KI010H03
            5', mRNA sequence.
ACCESSION   AV602537
VERSION     AV602537.1  GI:9724863
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE     Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
  MEDLINE   21570554
COMMENT     Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES              Location/Qualifiers
     source           1. .450
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone="E1KI010H03"
                      /clone_lib="Bos taurus kidney fetus"
                      /tissue_type="kidney"
                      /dev_stage="fetus"
                      /lab_host="DH10B"
                      /note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
                      was deleted from a NotI site"
BASE COUNT       91 a    131 c    135 g     92 t      1 others
ORIGIN


  Query Match             84.4%;  Score 60.8;  DB 9;  Length 450;
  Best Local Similarity   90.3%;  Pred. No. 1.1e-08;
  Matches   65;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||||||||||||||| |||||||||||| |||| |||||||||||||| ||||||
Db    136 ATGGCCTCTCTTGGCCTCCAGCTTGTGGGCTACGTCCTGGGCCTTCTGGGGCTGTTGGGC 195

Qy     61 acactggttgcc 72
          ||  || |||||
Db    196 ACCGTGATTGCC 207


RESULT   8
BE480970
LOCUS       BE480970                517 bp    mRNA    linear   EST 28-AUG-2000
DEFINITION  166244 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE480970
VERSION     BE480970.1  GI:9600503
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Sonstegard,T.S., Capuco,A.V., Van Tassell,C.P., Ashwell,M.S. and
            Wells,K.D.
  TITLE     Mapping of Expressed Sequence Tags from a normalized bovine mammary
            gland cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bdlg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@lpsi.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 18  row: I  column: 20
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .517
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="BARC 5BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled mRNA isolated from mammary
                      tissues at eight physiological, developmental, and disease
                      states."
BASE COUNT      108 a    144 c    158 g    107 t
ORIGIN


  Query Match             84.4%;  Score 60.8;  DB 10;  Length 517;
  Best Local Similarity   90.3%;  Pred. No. 1.2e-08;
  Matches   65;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;
```

```
Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||| |||||||||||| |||| |||||||||||||| ||||||
Db    150 ATGGCCTCTCTTGGCCTCCAGCTTGTGGGCTACGTCCTGGGCCTTCTGGGGCTGTTGGGC 209

Qy     61 acactggttgcc 72
          ||  || |||||
Db    210 ACCGTGATTGCC 221


RESULT  9
AV605969
LOCUS       AV605969                551 bp    mRNA    linear   EST 28-NOV-2001
DEFINITION  AV605969 Bos taurus kidney fetus Bos taurus cDNA clone E1KI032A05
            5', mRNA sequence.
ACCESSION   AV605969
VERSION     AV605969.1  GI:9736342
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE     Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
  MEDLINE   21570554
COMMENT     Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="E1KI032A05"
                     /clone_lib="Bos taurus kidney fetus"
                     /tissue_type="kidney"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /note="Vector: pZL1; Site_1: Sal1; Site_2: Not1; Poly A
                     was deleted from a Not1 site"
BASE COUNT      105 a    151 c    163 g    120 t     12 others
ORIGIN


  Query Match             84.4%;  Score 60.8;  DB 9;  Length 551;
  Best Local Similarity   90.3%;  Pred. No. 1.2e-08;
  Matches    65;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||| |||||||||||| |||| |||||||||||||| ||||||
Db    141 ATGGCCTCTCTTGGCCTCCAGCTTGTGGGCTACGTCCTGGGCCTTCTGGGGCTGTTGGGC 200

Qy     61 acactggttgcc 72
          ||  || |||||
Db    201 ACCGTGATTGCC 212


RESULT  10
AV604013
LOCUS       AV604013                555 bp    mRNA    linear   EST 28-NOV-2001
DEFINITION  AV604013 Bos taurus kidney fetus Bos taurus cDNA clone E1KI020B11
            5', mRNA sequence.
ACCESSION   AV604013
VERSION     AV604013.1  GI:9734386
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE     Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
  MEDLINE   21570554
COMMENT     Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="E1KI020B11"
                     /clone_lib="Bos taurus kidney fetus"
                     /tissue_type="kidney"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /note="Vector: pZL1; Site_1: Sal1; Site_2: Not1; Poly A
                     was deleted from a Not1 site"
BASE COUNT      106 a    158 c    167 g    124 t
ORIGIN


  Query Match             84.4%;  Score 60.8;  DB 9;  Length 555;
  Best Local Similarity   90.3%;  Pred. No. 1.2e-08;
  Matches    65;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||| |||||||||||| |||| |||||||||||||| ||||||
Db    137 ATGGCCTCTCTTGGCCTCCAGCTTGTGGGCTACGTCCTGGGCCTTCTGGGGCTGTTGGGC 196

Qy     61 acactggttgcc 72
          ||  || |||||
Db    197 ACCGTGATTGCC 208


RESULT  11
BE667438
LOCUS       BE667438                569 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  151799 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE667438
VERSION     BE667438.1  GI:10028029
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
```

```
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 64  row: M  column: 1
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 4BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
BASE COUNT      106 a    164 c    168 g    131 t
ORIGIN


  Query Match            84.4%;  Score 60.8;  DB 10;  Length 569;
  Best Local Similarity  90.3%;  Pred. No. 1.2e-08;
  Matches   65;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||||||||||||||| |||||||||||| |||| ||||||||||||| ||||||
Db    108 ATGGCCTCTCTTGGCCTCCAGCTTGTGGGCTACGTCCTGGGCCTTCTGGGGCTGTTGGGC 167

Qy     61 acactggttgcc 72
          ||  || |||||
Db    168 ACCGTGATTGCC 179


RESULT  12
BG325755
LOCUS       BG325755                728 bp    mRNA    linear   EST 27-FEB-2001
DEFINITION  602424466F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4562273 5',
            mRNA sequence.
ACCESSION   BG325755
VERSION     BG325755.1  GI:13132179
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1274  row: d  column: 18
            High quality sequence stop: 708.
FEATURES             Location/Qualifiers
     source          1. .728
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4562273"
                     /clone_lib="NIH_MGC_14"
                     /tissue_type="renal cell adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      154 a    205 c    183 g    186 t
ORIGIN


  Query Match            83.3%;  Score 60;  DB 10;  Length 728;
  Best Local Similarity  98.6%;  Pred. No. 2.3e-08;
  Matches   71;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     30 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGC-TTTGGGC 88

Qy     61 acactggttgcc 72
          ||||||||||||
Db     89 ACACTGGTTGCC 100


RESULT  13
BF541534
LOCUS       BF541534                856 bp    mRNA    linear   EST 11-DEC-2000
DEFINITION  602067906F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4067065 5',
            mRNA sequence.
ACCESSION   BF541534
VERSION     BF541534.1  GI:11628915
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 856)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM905  row: k  column: 02
            High quality sequence stop: 597.
FEATURES             Location/Qualifiers
     source          1. .856
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4067065"
                     /clone_lib="NIH_MGC_58"
                     /tissue_type="hypernephroma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.35 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
```

```
                    full-length clones and was constructed by Clontech
                    Laboratories (Palo Alto, CA)."
BASE COUNT      200 a    222 c    252 g    181 t      1 others
ORIGIN


  Query Match             83.3%;  Score 60;  DB 10;  Length 856;
  Best Local Similarity   98.6%;  Pred. No. 2.4e-08;
  Matches   71;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    138 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGC-TTTGGGC 196

Qy     61 acactggttgcc 72
          ||||||||||||
Db    197 ACACTGGTTGCC 208


RESULT  14
BF032123
LOCUS       BF032123                886 bp    mRNA    linear   EST 10-OCT-2000
DEFINITION  601559727F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3829254 5',
            mRNA sequence.
ACCESSION   BF032123
VERSION     BF032123.1  GI:10739835
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 886)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM501  row: f  column: 07
            High quality sequence stop: 584.
FEATURES             Location/Qualifiers
     source          1. .886
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3829254"
                     /clone_lib="NIH_MGC_58"
                     /tissue_type="hypernephroma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.35 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      199 a    219 c    262 g    206 t
ORIGIN


  Query Match             83.3%;  Score 60;  DB 10;  Length 886;
  Best Local Similarity   98.6%;  Pred. No. 2.4e-08;
  Matches   71;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    128 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGC-TTTGGGC 186

Qy     61 acactggttgcc 72
          ||||||||||||
Db    187 ACACTGGTTGCC 198


RESULT  15
BG164062
LOCUS       BG164062                978 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602341087F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4449212 5',
            mRNA sequence.
ACCESSION   BG164062
VERSION     BG164062.1  GI:12670765
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 978)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10232  row: m  column: 21
            High quality sequence stop: 689.
FEATURES             Location/Qualifiers
     source          1. .978
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4449212"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      220 a    279 c    276 g    203 t
ORIGIN


  Query Match             83.3%;  Score 60;  DB 10;  Length 978;
  Best Local Similarity   98.6%;  Pred. No. 2.5e-08;
  Matches   71;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     25 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGC-TTTGGGC 83

Qy     61 acactggttgcc 72
          ||||||||||||
Db     84 ACACTGGTTGCC 95


Search completed: June  2, 2002, 20:00:09
Job time: 14619 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 18:57:16 ; Search time 2378.87 Seconds
(without alignments)
6069.819 Million cell updates/sec

Title: US-09-663-600A-139_COPY_36_725
Perfect score: 690
Sequence: 1 atggcctctcttggcctcca.....acagcctgacagggtatgtg 690

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	690	100.0	693	9	AF250558	AF250558 Homo sapi
2	690	100.0	1475	6	AX092348	AX092348 Sequence
3	690	100.0	1475	6	AX299996	AX299996 Sequence
4	690	100.0	1506	9	BC014424	BC014424 Homo sapi
5	690	100.0	1918	9	AF177340	AF177340 Homo sapi
6	690	100.0	2863	6	AX136129	AX136129 Sequence
7	690	100.0	109465	9	AL158821	AL158821 Human DNA
8	689.6	99.9	1441	6	AX286822	AX286822 Sequence
9	562	81.4	953	4	AF358907	AF358907 Canis fam
10	546	79.1	779	6	AX286824	AX286824 Sequence
11	546	79.1	791	6	E31591	E31591 Tight junct
12	546	79.1	791	10	AF072128	AF072128 Mus muscu
13	546	79.1	2828	10	BC015252	BC015252 Mus muscu
14	361.6	52.4	615	6	AX136475	AX136475 Sequence
15	215.8	31.3	720	9	HSA132445	AJ132445 Homo sapi
16	215.8	31.3	1233	9	BC012126	BC012126 Homo sapi
17	215.8	31.3	1938	9	AF314090	AF314090 Homo sapi
c 18	215.8	31.3	100058	9	AP000695	AP000695 Homo sapi
c 19	215.8	31.3	101608	9	AP000694	AP000694 Homo sapi
20	215.8	31.3	120000	2	HSAC000005	AC000005 Homo sapi
21	215.8	31.3	175835	2	AC024074	AC024074 Homo sapi
c 22	215.8	31.3	340000	9	AP001726	AP001726 Homo sapi
23	212	30.7	49323	2	AC107455	AC107455 Mus muscu
24	206.2	29.9	1456	10	AF314089	AF314089 Mus muscu
25	203	29.4	729	10	AF124429	AF124429 Mus muscu
26	201.6	29.2	71409	5	AC091535	AC091535 Takifugu
27	172.2	25.0	167769	2	AC098891	AC098891 Rattus no
28	165	23.9	182	6	AX072760	AX072760 Sequence
29	155.6	22.6	1239	9	BC020838	BC020838 Homo sapi
30	155.6	22.6	116256	9	AL139101	AL139101 Human DNA
31	155.6	22.6	139248	2	AL353173	AL353173 Homo sapi
c 32	154.8	22.4	109982	10	AC091752	AC091752 Rattus no
33	152.6	22.1	1149	5	AF359426	AF359426 Danio rer
34	152.6	22.1	75015	5	AC091626	AC091626 Danio rer
35	150.6	21.8	633	10	AF087822	AF087822 Mus muscu
36	150.6	21.8	1783	10	AB000713	AB000713 Mus muscu
c 37	150.6	21.8	143322	10	AC079938	AC079938 Mus muscu
c 38	150	21.7	206358	2	AC098838	AC098838 Mus muscu
c 39	148.4	21.5	155195	2	AC092518	AC092518 Felis cat
40	145.2	21.0	194030	2	AC095021	AC095021 Sus scrof
41	144.2	20.9	755	5	BRE011789	AJ011789 Brachydan
42	144	20.9	1863	5	AF359432	AF359432 Danio rer
c 43	144	20.9	75015	5	AC091626	AC091626 Danio rer
44	142	20.6	1620	9	D88492	D88492 Cercopithec
45	141	20.4	681	6	AX276968	AX276968 Sequence

ALIGNMENTS

RESULT 1
AF250558
LOCUS AF250558 693 bp mRNA linear PRI 09-AUG-2000
DEFINITION Homo sapiens claudin-2 mRNA, complete cds.
ACCESSION AF250558
VERSION AF250558.1 GI:9755008
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 693)
AUTHORS Reinecker,H.-C., Sakaguchi,T. and Golden,H.M.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Gastrointestinal Unit, Massachusetts
General Hospital, Fruit Street, Boston, MA 02114, USA
FEATURES Location/Qualifiers
source 1. .693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq22.3-q23"
/cell_line="T-84"

```
                     /tissue_type="epithelium"
     CDS             1. .693
                     /note="tight junction-associated protein"
                     /codon_start=1
                     /product="claudin-2"
                     /protein_id="AAF98151.1"
                     /db_xref="GI:9755009"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                     RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                     SEFNSYSLTGYV"
BASE COUNT      142 a    202 c    168 g    181 t
ORIGIN


  Query Match             100.0%;  Score 690;  DB 9;  Length 693;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-191;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 60

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 120

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 180

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 240

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 300

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 360

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 420

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 480

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 540

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 600

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 660

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    661 TTCAATTCCTACAGCCTGACAGGGTATGTG 690


RESULT   2
AX092348
LOCUS       AX092348               1475 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 79 from Patent WO0116318.
ACCESSION   AX092348
VERSION     AX092348.1  GI:13444488
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1475)
  AUTHORS   Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
            Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
            Wood,W.I.
  TITLE     Secreted and transmembrane polypeptides and nucleic acids encoding
            the same
  JOURNAL   Patent: WO 0116318-A 79 08-MAR-2001;
            Genentech, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1475
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      345 a    404 c    376 g    350 t
ORIGIN


  Query Match             100.0%;  Score 690;  DB 6;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-191;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 181

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 241

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 301

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 361

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 421

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 481

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 541

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 601

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 661

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 721

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 781

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
```

```
Db    782 TTCAATTCCTACAGCCTGACAGGGTATGTG 811


RESULT   3
AX299996
LOCUS       AX299996               1475 bp    DNA     linear   PAT 26-NOV-2001
DEFINITION  Sequence 1 from Patent WO0166740.
ACCESSION   AX299996
VERSION     AX299996.1  GI:17129473
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Eaton,D.L., Fong,S., Goddard,A., Godowski,P.J., Grimaldi,C.J.,
            Gurney,A.L., Tumas,D., Watanabe,C.K., Wood,W.I. and Zhang,Z.
  TITLE     Compositions and methods for the treatment of immune related
            diseases
  JOURNAL   Patent: WO 0166740-A 1 13-SEP-2001;
            Genentech, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1475
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      345 a    404 c    376 g    350 t
ORIGIN


  Query Match             100.0%;  Score 690;  DB 6;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-191;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 181

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 241

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 301

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 361

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 421

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 481

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 541

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 601

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 661

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 721

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 781

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          ||||||||||||||||||||||||||||||
Db    782 TTCAATTCCTACAGCCTGACAGGGTATGTG 811


RESULT   4
BC014424
LOCUS       BC014424               1506 bp    mRNA    linear   PRI 19-SEP-2001
DEFINITION  Homo sapiens, Similar to claudin 2, clone MGC:20191 IMAGE:4645075,
            mRNA, complete cds.
ACCESSION   BC014424
VERSION     BC014424.1  GI:15680158
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1506)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 29 Row: m Column: 11.
FEATURES             Location/Qualifiers
     source          1. .1506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:20191 IMAGE:4645075"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_15"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             122. .814
                     /codon_start=1
                     /product="Similar to claudin 2"
                     /protein_id="AAH14424.1"
                     /db_xref="GI:15680159"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                     RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                     SEFNSYSLTGYV"
BASE COUNT      377 a    404 c    376 g    349 t
ORIGIN


  Query Match             100.0%;  Score 690;  DB 9;  Length 1506;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.5e-191;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 181

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 241

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 301

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 361

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 421

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 481

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 541

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 601

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 661

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 721

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 781

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
           |||||||||||||||||||||||||||||
Db     782 TTCAATTCCTACAGCCTGACAGGGTATGTG 811


RESULT   5
AF177340
LOCUS       AF177340               1918 bp    mRNA    linear   PRI 03-OCT-2000
DEFINITION  Homo sapiens clone SP82 claudin 2 mRNA, complete cds.
ACCESSION   AF177340
VERSION     AF177340.1  GI:10503979
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1918)
  AUTHORS   Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
            Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
            Yu,J. and Han,L.H.
  TITLE     Novel human cDNA clone with function of inhibiting cancer cell
            growth
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1918)
  AUTHORS   Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
            Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
            Yu,J. and Han,L.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1999) National Laboratory For Oncogenes & Related
            Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie Tu Road, Shanghai
            200032, P.R. China
FEATURES             Location/Qualifiers
     source          1. .1918
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="SP82"
     CDS             520. .1212
                     /note="similar to Mus musculus claudin 2"
                     /codon_start=1
                     /product="claudin 2"
                     /protein_id="AAG17984.1"
                     /db_xref="GI:10503980"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                     RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                     SEFNSYSLTGYV"
BASE COUNT      486 a    525 c    473 g    434 t
ORIGIN


  Query Match             100.0%;  Score 690;  DB 9;  Length 1918;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-191;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 579

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 639

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 699

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 759

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 819

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 879

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 939

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 999

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 1059

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 1119

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 1179
```

```
Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          ||||||||||||||||||||||||||||||
Db   1180 TTCAATTCCTACAGCCTGACAGGGTATGTG 1209


RESULT   6
AX136129
LOCUS       AX136129               2863 bp    DNA     linear   PAT 30-MAY-2001
DEFINITION  Sequence 51 from Patent EP1067182.
ACCESSION   AX136129
VERSION     AX136129.1  GI:14272537
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2863)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
            Hayashi,K.
  TITLE     Secretory protein or membrane protein
  JOURNAL   Patent: EP 1067182-A 51 10-JAN-2001;
            Helix Research Institute (JP)
FEATURES             Location/Qualifiers
     source          1. .2863
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             231. .923
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC39707.1"
                     /db_xref="GI:14272538"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                     RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                     SEFNSYSLTGYV"
BASE COUNT      670 a    794 c    711 g    688 t
ORIGIN


  Query Match           100.0%;  Score 690;  DB 6;  Length 2863;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-191;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 290

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 350

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 410

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 470

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 530

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 590

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 650

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 710

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    711 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 770

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    771 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 830

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    831 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 890

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          ||||||||||||||||||||||||||||||
Db    891 TTCAATTCCTACAGCCTGACAGGGTATGTG 920


RESULT   7
AL158821
LOCUS       AL158821               109465 bp    DNA     linear   PRI 08-FEB-2002
DEFINITION  Human DNA sequence from clone RP1-75H8 on chromosome Xq22.3-23
            Contains the CLDN2 gene for claudin 2, the gene for a novel protein
            similar to  KIAA0136, a eukaryotic translation elongation factor 1
            alpha 1 (EEF1A1) pseudogene and a CpG island, complete sequence.
ACCESSION   AL158821
VERSION     AL158821.16  GI:14329875
KEYWORDS    HTG; claudin; CLDN2; CpG island; EEF1A1; KIAA0136.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109465)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 8, 2001 this sequence version replaced gi:14160908.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP1-75H8 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone RP1-75H8
            It may be shorter because we sequence overlapping sections only
            once, except for a short overlap.
            The true left end of clone RP1-75H8 is at 1 in this sequence. The
            true left end of clone RP13-383K5 is at 109366 in this sequence.
FEATURES             Location/Qualifiers
```

```
source          1. .109465
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="X"
                /map="q22.3-23"
                /clone="RP1-75H8"
                /clone_lib="RPCI-1"
repeat_region   1. .62
                /note="Alu repeat: matches 1. .62 of consensus"
repeat_region   64. .1090
                /note="L2 repeat: matches 1154. .2194 of consensus"
repeat_region   1070. .1202
                /note="L2 repeat: matches 670. .806 of consensus"
repeat_region   1558. .1615
                /note="29 copies 2 mer ct 69% conserved"
repeat_region   1732. .1858
                /note="L2 repeat: matches 2579. .2710 of consensus"
repeat_region   1901. .2047
                /note="MIR repeat: matches 109. .262 of consensus"
repeat_region   2112. .2514
                /note="L2 repeat: matches 2095. .2477 of consensus"
repeat_region   2515. .2806
                /note="AluSq repeat: matches 1. .293 of consensus"
repeat_region   2807. .3358
                /note="L2 repeat: matches 1485. .2095 of consensus"
repeat_region   3374. .3580
                /note="AluJb repeat: matches 4. .199 of consensus"
repeat_region   3581. .3610
                /note="15 copies 2 mer ac 100% conserved"
repeat_region   3617. .3684
                /note="FRAM/FAM repeat: matches -4. .63 of consensus"
repeat_region   3685. .3716
                /note="16 copies 2 mer ac 100% conserved"
repeat_region   3717. .3891
                /note="AluJb repeat: matches 127. .301 of consensus"
mRNA            4841. .6387
                /gene="CLDN2"
                /note="match: cDNAs: Em:AF250558 Em:AK004990 Em:AF358908
                Em:AF177340 Em:AB000713 Em:BC014424 Em:BC015252
                Em:AK020255 Em:AF314089 Em:AF087822 Em:AJ011789
                Em:AF358907 Em:AF314090 Em:AF072128 Em:BC012126 Em:D88492
                match: ESTs: Em:AW107793 Em:AA345449 Em:BG164241
                Em:BG423208 Em:T86266 Em:BE304667 Em:BG951731 Em:AA639438
                Em:BE480970 Em:BG951371 Em:BG328625 Em:BG325755
                Em:BG951374 Em:AI790813 Em:AW006504 Em:BF819923
                Em:AI262767 Em:BB847585 Em:BI006997 Em:BG693582
                Em:AI685439 Em:AI949846 Em:BB845010 Em:BG985977 Em:T86158
                Em:BG469035 Em:AA976345 Em:AV604013 Em:AI659323
                Em:BF104997 Em:BB653578 Em:BG385562 Em:AI913916
                Em:BE048787 Em:AW083920 Em:AV647382 Em:AI685788
                Em:BI010035 Em:BI006211 Em:AW837723 Em:AW837724
                Em:AW385836 Em:AW837727 Em:BE513091 Em:BG423109
                Em:BG164062 Em:AI285131 Em:BF541534 Em:AV605969
                Em:AI953481 Em:AI343787 Em:AI769564 Em:AW194732
                Em:BF032123 Em:AW015998 Em:BG426062 Em:AA552647
                Em:BB848306 Em:AI510548 Em:BF758311 Em:BG386074
                Em:BI143586 Em:BG399901 Em:BI008886 Em:AI932964
                Em:AV602536 Em:AV602537"
                /product="dJ75H8.1 (claudin 2)"
                /evidence=not_experimental
gene            4841. .6387
                /gene="CLDN2"
CDS             5019. .5711
                /gene="CLDN2"
                /note="match: proteins: Sw:O88552"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ75H8.1 (claudin 2)"
                /protein_id="CAD23055.1"
                /db_xref="GI:18643740"
                /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV
                GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGM
                RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK
                SEFNSYSLTGYV"
polyA_signal    6369. .6374
                /gene="CLDN2"
polyA_site      6387
                /gene="CLDN2"
repeat_region   7257. .7654
                /note="L2 repeat: matches 2322. .2748 of consensus"
repeat_region   7895. .8027
                /note="MIR repeat: matches 76. .201 of consensus"
repeat_region   9628. .9756
                /note="L2 repeat: matches 2566. .2693 of consensus"
repeat_region   10976. .11272
                /note="AluY repeat: matches 1. .299 of consensus"
repeat_region   11462. .11792
                /note="L1PA2 repeat: matches 5821. .6146 of consensus"
repeat_region   12308. .12478
                /note="MIR repeat: matches 2. .179 of consensus"
repeat_region   13066. .13208
                /note="L2 repeat: matches 2359. .2491 of consensus"
repeat_region   14358. .14430
                /note="L2 repeat: matches 2664. .2732 of consensus"
repeat_region   14441. .14555
                /note="L2 repeat: matches 2574. .2692 of consensus"
repeat_region   14511. .14633
                /note="L2 repeat: matches 2570. .2705 of consensus"
repeat_region   14633. .15490
                /note="L2 repeat: matches 1702. .2606 of consensus"
repeat_region   15618. .16068
                /note="L2 repeat: matches 1967. .2418 of consensus"
mRNA            complement(join(17526. .18422,18728. .19011,19305. .19995,
                31703. .31878,33270. .33324,33726. .33794,35126. .35254,
                38802. .38900,53374. .53474,54870. .54989,57681. .57809,
                58135. .58267,61886. .62032,62772. .62990,70020. .70152,
                76467. .76539,76658. .77002))
                /gene="dJ75H8.2"
                /note="The splice sites between exon 4 and 5 are not
                canonical.
                match: cDNAs: Em:AK017472 Em:AK056235 Em:D50926
                Em:AK021627
                match: ESTs: Em:BG962578 Em:BF161943 Em:AW880880
                Em:BG977099 Em:AW881141 Em:AW881142 Em:BB373111
                Em:BB665619 Em:BB649318"
                /product="dJ75H8.2 (similar to KIAA0136 protein)"
                /evidence=not_experimental
gene            complement(17526. .77002)
                /gene="dJ75H8.2"
polyA_site      complement(17526)
                /gene="dJ75H8.2"
polyA_signal    complement(17543. .17548)
                /gene="dJ75H8.2"
repeat_region   18154. .18201
                /note="24 copies 2 mer ga 93% conserved"
CDS             complement(join(18269. .18422,18728. .19011,19305. .19995,
                31703. .31878,33270. .33324,33726. .33794,35126. .35254,
                38802. .38900,53374. .53474,54870. .54989,57681. .57809,
                58135. .58267,61886. .62032,62772. .62990,70020. .70152,
                76467. .76539))
                /gene="dJ75H8.2"
                /note="match: proteins: Sw:Q14149"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ75H8.2 (similar to KIAA0136 protein)"
                /protein_id="CAD23056.1"
                /db_xref="GI:18643741"
                /translation="MSPRYLQSNSSSHTRPFSAIAELLDNAVDPDVSARTVFIDVEEV
                KNKSCLTFTDDGCGMTPHKLHRMLSFGFTDKVIKKSQCPIGVFGNGFKSGSMRLGKDA
                LVFTKNGGTLTVGLLSQTYLECVQAQAVIVPIVPFNQQNKKMIITEDSLPSLEAILNY
                SIFNRENDLLAQFDAIPGKKGTRVLIWNIRRNKNGKSELDFDTDQYDILVSDFDTEEK
                MTGGVTSELPETEYSLRAFCGILYMKPRMKIFLRQKKVTTQMIAKSLANVEYDTYKPT
                FTNKQVRITFGFSCKNSNQFGIMMYHNNRLIKSFEKVGCQVKPTRGEGVGVIGVIECN
                FLKPAYNKQDFEYTKEYRLTINALAQKLNAYWKEKTSQDNFETSTVARPIPKVPDQTW
                VQCDECLKWRKLPGKIDPSMLPARWFCYYNSHPKYRRCSVPEEQELTDEDLCLSKAKK
```

```
                     QEQTVEEKKKMPMENENHQVFSNPPKILTVQEMAGLNNKTIGYEGIHSPSVLPSGGEE
                     SRSPSLQLKPLDSSVLQFSSKYKWILGEEPVEKRRRLQNEMTTPSLDYSMPAPYRRVE
                     APVAYPEGENSHDKSSSERSTPPYLFPEYPEASKNTGQNREVSILYPGAKDQRQGSLL

  Query Match           100.0%;  Score 690;  DB 9;  Length 109465;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-191;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5019 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 5078

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5079 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 5138

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5139 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 5198

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5199 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 5258

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5259 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 5318

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5319 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 5378

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5379 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 5438

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5439 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 5498

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5499 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 5558

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5559 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 5618

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5619 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 5678

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db   5679 TTCAATTCCTACAGCCTGACAGGGTATGTG 5708


RESULT   8
AX286822
LOCUS       AX286822               1441 bp    DNA     linear   PAT 21-NOV-2001
DEFINITION  Sequence 1 from Patent WO0180879.
ACCESSION   AX286822
VERSION     AX286822.1  GI:17048854
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vinals y de Bassols,C.G., Gaulis,S.R., Coche,T.G. and Orntoft,T.A.
  TITLE     Colorectal cancer vaccines and diagnosis
  JOURNAL   Patent: WO 0180879-A 1 01-NOV-2001;
            SMITHKLINE,BEECHAM BIOLOGICALS (S.A.)

FEATURES             Location/Qualifiers
     source          1. .1441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      337 a    399 c    357 g    345 t      3 others
ORIGIN


  Query Match           99.9%;  Score 689.6;  DB 6;  Length 1441;
  Best Local Similarity  99.9%;  Pred. No. 2e-191;
  Matches  689;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 136

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 196

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 256

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 316

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 376

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 436

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 496

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 556

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 616

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db    617 CTCTGCTTTTCCTGCTCATSCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 676

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 736

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    737 TTCAATTCCTACAGCCTGACAGGGTATGTG 766


RESULT   9
AF358907
LOCUS       AF358907               953 bp    mRNA    linear   MAM 08-MAY-2001
DEFINITION  Canis familiaris integral membrane protein claudin-2 mRNA, complete
            cds.
ACCESSION   AF358907
VERSION     AF358907.1  GI:13991612
KEYWORDS    .
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
```

```
REFERENCE   1  (bases 1 to 953)
  AUTHORS   Furuse,M., Furuse,K., Sasaki,H. and Tsukita,S.
  TITLE     Conversion of Zonulae Occludentes from Tight to Leaky Strand Type
            by Introducing Claudin-2 into Madin-Darby Canine Kidney I Cells
  JOURNAL   J. Cell Biol. 153 (2), 263-272 (2001)
  MEDLINE   21206012
   PUBMED   11309408
REFERENCE   2  (bases 1 to 953)
  AUTHORS   Furuse,M., Furuse,K., Sasaki,H. and Tsukita,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2001) Cell Biology, Kyoto University, Faculty of
            Medicine, Yoshida-Konoe, Sakyo, Kyoto 606-8501, Japan
FEATURES             Location/Qualifiers
     source          1. .953
                     /organism="Canis familiaris"
                     /db_xref="taxon:9615"
                     /sex="male"
                     /tissue_type="liver"
     CDS             59. .751
                     /note="localizes to tight junctions"
                     /codon_start=1
                     /product="integral membrane protein claudin-2"
                     /protein_id="AAK51433.1"
                     /db_xref="GI:13991613"
                     /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWRTSSYVGTSIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIVSVVGM
                     RCTVFCQDSRAKDRLAVVGGVFFIIGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISSLFSLVAGIILCFSCPLQGNRSDYYDSYQAQPLATRGSPRPGQPPKAK
                     SEFNSYSLTGYV"
BASE COUNT      195 a    289 c    265 g    204 t
ORIGIN
```

```
  Query Match             81.4%;  Score 562;  DB 4;  Length 953;
  Best Local Similarity   88.4%;  Pred. No. 5.8e-154;
  Matches  610;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||||||| |||||||||||||| |||||||||||||||||| |||||||| ||||||
Db      59 ATGGCCTCTCTCGGCCTCCAACTTGTAGGCTACATCCTAGGCCTCCTGGGGCTGTTGGGC 118

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           || |||||| |||||||||| |||||||||  |||||| || || || ||| ||||||||
Db     119 ACCCTGGTGGCCATGCTGCTTCCCAGCTGGCGAACAAGCTCCTACGTTGGTACCAGCATC 178

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||| || || ||||||||||||||||||||||||||| || ||||||||||||||||||
Db     179 GTGACGGCGGTCGGCTTCTCCAAGGGCCTCTGGATGGAGTGCGCCACACACAGCACAGGC 238

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           || |||||||||||||||| |||||||| || |||||||| ||||||||||||||||||||
Db     239 ATAACCCAGTGTGACATCTACAGCACCCTCCTAGGCCTGCCTGCTGACATCCAGGCTGCC 298

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||| ||||| ||  |||||||||||| ||||||||||||
Db     299 CAGGCCATGATGGTGACATCCAGTGCGATCTCTTCGTTGGCCTGCATTGTCTCTGTGGTG 358

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           |||||||||||||| |||||||||||||| |||||||||||||||||| ||||||| |
Db     359 GGCATGAGATGCACTGTCTTCTGCCAGGACTCCCGAGCCAAAGACAGACTGGCGGTGGTG 418

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           |||||||||| |||||| ||||||||||||||| ||||| || |||||||||| |||
Db     419 GGTGGAGTCTTCTTCATCATTGGAGGCCTCCTGGGCTTCATCCCCGTTGCCTGGAACCTT 478

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           || |||||||| |||||||||||||| || ||||| || || |||||||| || ||||||
Db     479 CACGGGATCCTGCGGGACTTCTACTCCCCGCTGGTACCCGATAGCATGAAGTTCGAGATC 538

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           ||||| ||||| ||| |||||||||||||| ||| |||||||||| ||||||| ||||||
Db     539 GGAGAAGCTCTCTACCTGGGCATTATTTCCTCCTTGTTCTCCCTGGTAGCTGGCATCATC 598

Qy     541 ctctgctttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           ||||||||||||||| ||  |||| |||||||||| ||||||| ||  ||||||| |||
Db     599 CTCTGCTTTTCCTGCCCACTCCAGGGAAATCGCTCCGACTACTATGACTCCTACCAGGCC 658

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
           || || ||||| || ||  |||||||||||||| |||||||| ||||||| |||||| |||
Db     659 CAGCCCCTTGCAACTAGAGGCTCTCCAAGGCCGGGTCAACCGCCCAAAGCCAAGAGCGAG 718

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
           || || ||||||||||||||||||||||||
Db     719 TTTAACTCCTACAGCCTGACAGGGTATGTG 748
```

```
RESULT  10
AX286824
LOCUS       AX286824                 779 bp    DNA     linear   PAT 21-NOV-2001
DEFINITION  Sequence 3 from Patent WO0180879.
ACCESSION   AX286824
VERSION     AX286824.1  GI:17048855
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Vinals y de Bassols,C.G., Gaulis,S.R., Coche,T.G. and Orntoft,T.A.
  TITLE     Colorectal cancer vaccines and diagnosis
  JOURNAL   Patent: WO 0180879-A 3 01-NOV-2001;
            SMITHKLINE BEECHAM BIOLOGICALS (S.A.)
FEATURES             Location/Qualifiers
     source          1. .779
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
BASE COUNT      152 a    205 c    216 g    206 t
ORIGIN
```

```
  Query Match             79.1%;  Score 546;  DB 6;  Length 779;
  Best Local Similarity   87.0%;  Pred. No. 2.9e-149;
  Matches  600;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           |||||||| |||||| | ||||| ||||||||||||||||||||||| ||||||| || |||
Db      48 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 107

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db     108 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 167

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||| || |||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db     168 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 227

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           |||||||||| || ||||| || |||||| | || || || ||||||||||||||||||||
Db     228 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 287

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db     288 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 347

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db     348 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 407

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           |||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db     408 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 467

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||| |||| |||||||||||||| || |||||||||||||||||||||||||||||||
Db     468 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGAGATT 527
```

```
Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          |||||||||||| |||||||||||  ||  ||  |||||||  ||   ||  ||||  |||  |||||
Db    528 GGAGAGGCTCTGTACTTGGGCATCATCTCAGCCCTGTTTTCTTTGGTAGCCGGAGTCATC 587

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||  ||||||||||||||  |||||  |  |||||  ||||||||||  ||||  ||||||  |||
Db    588 CTTTGCTTTTCCTGCTCGCCCCAGGGCAATCGTACCAACTACTATGATGGCTACCAGGCC 647

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||  ||||||||||| ||||||||||||||  |||  |||||  |||||||  ||||||||||
Db    648 CAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGCTCAACAGCCCAAAGCCAAGAGTGAG 707

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||  ||  ||||||||||| |||||||||
Db    708 TTCAACTCATACAGCCTGACTGGGTATGTG 737


RESULT  11
E31591
LOCUS       E31591                   791 bp    DNA     linear   PAT 07-FEB-2001
DEFINITION  Tight junction-constituting membrane protein claudin family.
ACCESSION   E31591
VERSION     E31591.1  GI:13018519
KEYWORDS    JP 2000032984-A/2.
SOURCE      Mus sp. .
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Syoichiro,T.
  TITLE     Tight junction-constituting membrane protein claudin family
  JOURNAL   Patent: JP 2000032984-A 2 02-FEB-2000;
            EISAI CO LTD
COMMENT     OS   Mus sp. (mouse)
            PN   JP 2000032984-A/2
            PD   02-FEB-2000
            PF   26-JUN-1998 JP 1998179847
            PR
            PI   SYOICHIRO TSUKITA
            PC   C12N15/09,C07K14/47,C07K16/18,C12N5/10,C12P21/02,C12Q1/02, PC
             C12Q1/68,
            PC   G01N33/15,G01N33/53,G01N33/577//C12P21/08,(C12N5/10,C12R1:91),
            PC   (C12P21/02,C12R1:91),C12N15/00,C12N5/00,(C12N5/00,C12R1:91) CC
              Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             48. .740.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
BASE COUNT      155 a    207 c    220 g    209 t
ORIGIN


  Query Match             79.1%;  Score 546;  DB 6;  Length 791;
  Best Local Similarity   87.0%;  Pred. No. 2.9e-149;
  Matches  600;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||  ||||||  |  |||||  ||||||||||||||||||||  |||||||  ||  |||
Db     48 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 107

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    ||||||||||||  ||||  ||||   |||  |||||  |||||  ||||||||||||
Db    108 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 167

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          |||||  ||  ||||||||  ||||||||||||||||||||  |||||  ||||||||||||||||
Db    168 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 227

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||| ||  |||||  ||  ||||||  |  ||  ||  ||  ||||||||||||||||||
Db    228 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 287

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||  ||||||||||  |||||  |||||  ||  ||||||||||||||||
Db    288 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 347

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||  ||  ||||||||||  ||  |||||  ||  ||||||||||||  ||||
Db    348 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 407

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||  |||  |||||||  ||  ||  ||  |||||  |||||||||
Db    408 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 467

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          |||||  |||||  ||||||||||||||  ||  |||||  ||||||||||||||||||||||||
Db    468 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGAGATT 527

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||| ||||||||||  ||  ||  |||||||  ||   ||  ||||  |||  |||||
Db    528 GGAGAGGCTCTGTACTTGGGCATCATCTCAGCCCTGTTTTCTTTGGTAGCCGGAGTCATC 587

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||  ||||||||||||||  |||||  |  |||||  ||||||||||  ||||  ||||||  |||
Db    588 CTTTGCTTTTCCTGCTCGCCCCAGGGCAATCGTACCAACTACTATGATGGCTACCAGGCC 647

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||  ||||||||||| ||||||||||||||  |||  |||||  |||||||  ||||||||||
Db    648 CAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGCTCAACAGCCCAAAGCCAAGAGTGAG 707

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||  ||  ||||||||||| |||||||||
Db    708 TTCAACTCATACAGCCTGACTGGGTATGTG 737


RESULT  12
AF072128
LOCUS       AF072128                 791 bp    mRNA    linear   ROD 21-JUL-1998
DEFINITION  Mus musculus claudin-2 mRNA, complete cds.
ACCESSION   AF072128
VERSION     AF072128.1  GI:3335183
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Furuse,M., Fujita,K., Hiiragi,T., Fujimoto,K. and Tsukita,S.
  TITLE     Claudin-1 and -2: novel integral membrane proteins localizing at
            tight junctions with No sequence similarity to occludin
  JOURNAL   J. Cell Biol. 141 (7), 1539-1550 (1998)
  MEDLINE   98311639
REFERENCE   2  (bases 1 to 791)
  AUTHORS   Furuse,M., Fujita,K., Hiiragi,T., Fujimoto,K. and Tsukita,S.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-1998) Cell Biology, Kyoto University Faculty of
            Medicine, Sakyo-ku, Kyoto 606-8501, Japan
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             48. .740
                     /note="integral membrane protein localizing at tight
                     junctions"
                     /codon_start=1
                     /product="claudin-2"
                     /protein_id="AAC27079.1"
                     /db_xref="GI:3335184"
                     /translation="MASLGVQLVGYILGLLGLLGTSIAMLLPNWRTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAMSSLACIISVVGM
                     RCTVFCQDSRAKDRVAVVGGVFFILGGILGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISALFSLVAGVILCFSCSPQGNRTNYYDGYQAQPLATRSSPRSAQQPKAK
                     SEFNSYSLTGYV"
```

```
BASE COUNT      155 a    207 c    220 g    209 t
ORIGIN

  Query Match             79.1%;  Score 546;  DB 10;  Length 791;
  Best Local Similarity   87.0%;  Pred. No. 2.9e-149;
  Matches  600;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db     48 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 107

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db    108 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 167

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db    168 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 227

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db    228 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 287

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          |||||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db    288 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 347

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db    348 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 407

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db    408 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 467

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||| ||||| |||||||||||||| || ||||| |||||||||||||||||||||||||
Db    468 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGAGATT 527

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          |||||||||||| |||||||||| || ||  ||||||| ||  || |||| ||| |||||
Db    528 GGAGAGGCTCTGTACTTGGGCATCATCTCAGCCCTGTTTTCTTTGGTAGCCGGAGTCATC 587

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          || |||||||||||||  ||||| | |||||  |||||||||| |||| |||||| |||
Db    588 CTTTGCTTTTCCTGCTCGCCCCAGGGCAATCGTACCAACTACTATGATGGCTACCAGGCC 647

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          || |||||||||| |||||||||||||  ||| |||||  ||||||| ||||||||||
Db    648 CAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGCTCAACAGCCCAAAGCCAAGAGTGAG 707

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          ||||| || |||||||||| |||||||||
Db    708 TTCAACTCATACAGCCTGACTGGGTATGTG 737


RESULT  13
BC015252
LOCUS       BC015252               2828 bp    mRNA    linear   ROD 29-OCT-2001
DEFINITION  Mus musculus, claudin 2, clone MGC:18348 IMAGE:4223446, mRNA,
            complete cds.
ACCESSION   BC015252
VERSION     BC015252.1  GI:15929633
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2828)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 25 Row: h Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7710003.
FEATURES             Location/Qualifiers
     source          1. .2828
                     /organism="Mus musculus"
                     /db_xref="LocusID:12738"
                     /db_xref="taxon:10090"
                     /clone="MGC:18348 IMAGE:4223446"
                     /tissue_type="Kidney, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             126. .818
                     /codon_start=1
                     /product="claudin 2"
                     /protein_id="AAH15252.1"
                     /db_xref="GI:15929634"
                     /translation="MASLGVQLVGYILGLLGLLGTSIAMLLPNWRTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAMSSLACIISVVGM
                     RCTVFCQDSRAKDRVAVVGGVFFILGGILGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISALFSLVAGVILCFSCSPQGNRTNYYDGYQAQPLATRSSPRSAQQPKAK
                     SEFNSYSLTGYV"
BASE COUNT      736 a    697 c    667 g    728 t
ORIGIN

  Query Match             79.1%;  Score 546;  DB 10;  Length 2828;
  Best Local Similarity   87.0%;  Pred. No. 3.1e-149;
  Matches  600;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db    126 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 185

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db    186 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 245

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db    246 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 305

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db    306 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 365

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          |||||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db    366 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 425

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db    426 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 485
```

```
Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db    486 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 545

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||| ||||| |||||||||||||| || ||||| ||||||||||||||||||||||||
Db    546 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGAGATT 605

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          |||||||||||| |||||||||| || ||  ||||||| ||  || |||| ||| |||||
Db    606 GGAGAGGCTCTGTACTTGGGCATCATCTCAGCCCTGTTTTCTTTGGTAGCCGGAGTCATC 665

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          || |||||||||||||  ||||| | |||||  |||||||||| |||| |||||| |||
Db    666 CTTTGCTTTTCCTGCTCGCCCCAGGGCAATCGTACCAACTACTATGATGGCTACCAGGCC 725

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          || |||||||||||  |||||||||||||||  ||| |||||  ||||||| ||||||||||
Db    726 CAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGCTCAACAGCCCAAAGCCAAGAGTGAG 785

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          ||||| || |||||||||||  |||||||||
Db    786 TTCAACTCATACAGCCTGACTGGGTATGTG 815


RESULT  14
AX136475
LOCUS       AX136475                615 bp    DNA     linear   PAT 30-MAY-2001
DEFINITION  Sequence 397 from Patent EP1067182.
ACCESSION   AX136475
VERSION     AX136475.1  GI:14272879
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
            Hayashi,K.
  TITLE     Secretory protein or membrane protein
  JOURNAL   Patent: EP 1067182-A 397 10-JAN-2001;
            Helix Research Institute (JP)
FEATURES             Location/Qualifiers
     source          1. .615
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      133 a    174 c    157 g    143 t      8 others
ORIGIN


  Query Match           52.4%;  Score 361.6;  DB 6;  Length 615;
  Best Local Similarity 97.4%;  Pred. No. 4.2e-95;
  Matches  375;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 290

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 350

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 410

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    411 ATCACCCAGTGTGACATCTATAACACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 470

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 CANGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 530

Qy    301 ggcatgagatgcacagtcttctgccagg-aatcccgagccaaagacagagtggcggtagc 359
          ||||||| |||||||||||||||||||| |||||||| ||||| ||| | |||||||| |
Db    531 GGCATGANATGCACAGTCTTCTGCCAGGNAATCCCGAACCAAANACANAATGGCGGTANC 590

Qy    360 aggtggagtctttttcatccttgga 384
          ||||||| ||||||||||||||||||
Db    591 AGGTGGAATCTTTTTCATCCTTGGA 615


RESULT  15
HSA132445
LOCUS       HSA132445               720 bp    DNA     linear   PRI 22-JAN-1999
DEFINITION  Homo sapiens CLDN14 gene.
ACCESSION   AJ132445
VERSION     AJ132445.1  GI:4186039
KEYWORDS    claudin-14; CLDN14 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Keen,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1999) Keen T.J., Molecular Medicine Unit,
            University of Leeds, Clinical Sciences Building, St James's
            Hospital, Leeds, LS9 7TF, UK
REFERENCE   2  (bases 1 to 720)
  AUTHORS   Keen,T.J. and Inglehearn,C.F.
  JOURNAL   Unpublished
COMMENT     Related sequence AC000005.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
     gene            1. .720
                     /gene="CLDN14"
     CDS             1. .720
                     /gene="CLDN14"
                     /codon_start=1
                     /product="claudin-14"
                     /protein_id="CAA10669.1"
                     /db_xref="GI:4186040"
                     /translation="MASTAVQLLGFLLSFLGMVGTLITTILPHWRRTAHVGTNILTAV
                     SYLKGLWMECVWHSTGIYQCQIYRSLLALPQDLQAARALMVISCLLSGIACACAVIGM
                     KCTRCAKGTPAKTTFAILGGTLFILAGLLCMVAVSWTTNDVVQNFYNPLLPSGMKFEI
                     GQALYLGFISSSLSLIGGTLLCLSCQDEAPYRPYQAPPRATTTTANTAPAYQPPAAYK
                     DNRAPSVTSATHSGYRLNDYV"
BASE COUNT      126 a    270 c    193 g    131 t
ORIGIN


  Query Match           31.3%;  Score 215.8;  DB 9;  Length 720;
  Best Local Similarity 61.8%;  Pred. No. 2.8e-52;
  Matches  343;  Conservative    0;  Mismatches  212;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||      | | | || ||| |||||| | | ||  || | |||||  |  |||||
Db      1 ATGGCCAGCACGGCCGTGCAGCTTCTGGGCTTCCTGCTCAGCTTCCTGGGCATGGTGGGC 60

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||  || |  |||   | || ||   ||||   |  |   |  | || ||  ||| |||
Db     61 ACGTTGATCACCACCATCCTGCCGCACTGGCGGAGGACAGCGCACGTGGGCACCAACATC 120

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          | || || ||   || |   || || ||||||||||| ||||     ||||||||||||
Db    121 CTCACGGCCGTGTCCTACCTGAAAGGGCTCTGGATGGAGTGTGTGTGGCACAGCACAGGC 180

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||  ||||||  | |||||  |  |||| ||||  ||||||   ||| ||||||||||||
Db    181 ATCTACCAGTGCCAGATCTACCGATCCCTGCTGGCGCTGCCCCAAGACCTCCAGGCTGCC 240
```

```
Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          |  ||| | ||||| |  ||| |     ||||   | | ||||||     | | ||  |
Db    241 CGCGCCCTCATGGTCATCTCCTGCCTGCTCTCGGGCATAGCCTGCGCCTGCGCCGTCATC 300

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          || ||||  |||||   || |  | |||   | |  |||||   ||   | ||  |
Db    301 GGGATGAAGTGCACGCGCTGCGCCAAGGGCACACCCGCCAAGACCACCTTTGCCATCCTC 360

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          || ||   | | |||||||| |  |||||||| |  |  |  | ||  ||||||
Db    361 GGCGGCACCCTCTTCATCCTGGCCGGCCTCCTGTGCATGGTGGCCGTCTCCTGGACCACC 420

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          | |   |  | | | ||||||||   || ||| ||||   | |||||||| |||||||||
Db    421 AACGACGTGGTGCAGAACTTCTACAACCCGCTGCTGCCCAGCGGCATGAAGTTTGAGATT 480

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||  |||| || ||| ||||| | || || ||     |||| || || | ||| | | |
Db    481 GGCCAGGCCCTGTACCTGGGCTTCATCTCCTCGTCCCTCTCGCTCATTGGTGGCACCCTG 540

Qy    541 ctctgcttttcctgc 555
          || ||| | ||||||
Db    541 CTTTGCCTGTCCTGC 555
```

Search completed: June 2, 2002, 18:58:05
Job time: 11105 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 15:56:30 ; Search time 3664.46 Seconds
(without alignments)
88.397 Million cell updates/sec

Title: US-09-663-600A-2_COPY_1_24
Perfect score: 24
Sequence: 1 gcauccuacucccauccaauucca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	19.2	80.0	438	12	AQ887612	AQ887612 HS_5557_A
	2	18.8	78.3	379	9	AV558212	AV558212 AV558212
	3	18.8	78.3	434	12	BH172929	BH172929 SALK_0063
	4	18.8	78.3	661	12	AG076330	AG076330 Pan trogl
	5	18.4	76.7	286	9	AV329061	AV329061 AV329061
c	6	18.4	76.7	361	9	AI866251	AI866251 wl27f06.x
	7	18.2	75.8	202	12	AQ248820	AQ248820 T16N9-Sp6
	8	18.2	75.8	284	10	BM089677	BM089677 503497 MA
c	9	18.2	75.8	321	9	AA153510	AA153510 mq57b05.r
	10	18.2	75.8	327	12	AZ442220	AZ442220 1M0234E19
	11	18.2	75.8	343	9	AW789229	AW789229 C00996-R
	12	18.2	75.8	411	10	BF601807	BF601807 266877 MA
c	13	18.2	75.8	431	9	AI645377	AI645377 mq57b05.y
c	14	18.2	75.8	462	9	BB860142	BB860142 BB860142
c	15	18.2	75.8	511	12	AQ978003	AQ978003 RPCI-23-3
	16	18.2	75.8	547	9	AW669920	AW669920 113680 MA
	17	18.2	75.8	577	12	BH033134	BH033134 RPCI-24-3
	18	18.2	75.8	605	10	BG689078	BG689078 336956 BA
c	19	18.2	75.8	637	9	BB635865	BB635865 BB635865
c	20	18.2	75.8	666	9	BB208585	BB208585 BB208585
c	21	18.2	75.8	703	9	AI048533	AI048533 ud61a09.y
	22	18.2	75.8	816	12	AG146561	AG146561 Pan trogl
c	23	17.8	74.2	297	10	BE352981	BE352981 DG1_2_F08
c	24	17.8	74.2	428	10	W43882	W43882 23275 CD4-1
	25	17.8	74.2	990	10	BG854007	BG854007 1024038G0
c	26	17.6	73.3	250	9	AV021672	AV021672 AV021672
	27	17.6	73.3	260	9	AW428024	AW428024 64802 MAR
c	28	17.6	73.3	265	12	AZ920824	AZ920824 1006021D0
	29	17.6	73.3	294	10	C20172	'C20172 C20172 Rice
	30	17.6	73.3	360	12	CNS01ZDQ	AL174167 Tetraodon
c	31	17.6	73.3	372	10	BI055852	BI055852 CM3-GN033
	32	17.6	73.3	401	10	BF991907	BF991907 QV3-GN020
c	33	17.6	73.3	419	12	BH031217	BH031217 RPCI-24-3
	34	17.6	73.3	429	9	AW484974	AW484974 63291 MAR
	35	17.6	73.3	502	10	BM482975	BM482975 536085 MA
c	36	17.6	73.3	507	10	BE405827	BE405827 WHE0437_B
	37	17.6	73.3	520	9	BE020754	BE020754 sm52d12.y
	38	17.6	73.3	567	10	BF231030	BF231030 253231 BA
c	39	17.6	73.3	575	10	BJ073394	BJ073394 BJ073394
	40	17.6	73.3	629	12	AQ782163	AQ782163 HS_3176_B
c	41	17.6	73.3	643	10	BJ074708	BJ074708 BJ074708
c	42	17.6	73.3	666	10	BJ073088	BJ073088 BJ073088
	43	17.6	73.3	673	12	AZ314706	AZ314706 1M0031F10
c	44	17.6	73.3	679	10	BJ064086	BJ064086 BJ064086
	45	17.6	73.3	685	10	BE254962	BE254962 601115223

ALIGNMENTS

RESULT 1
AQ887612/c
LOCUS AQ887612 438 bp DNA linear GSS 10-NOV-1999
DEFINITION HS_5557_A2_G10_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=9325 Col=20 Row=M, DNA sequence.
ACCESSION AQ887612
VERSION AQ887612.1 GI:6343802
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 438)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 9325 row: M column: 20
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 438.
FEATURES Location/Qualifiers
source 1. .438
/organism="Homo sapiens"

```
                     /db_xref="taxon:9606"
                     /clone="Plate=9325 Col=20 Row=M"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      154 a     54 c     67 g    162 t      1 others
ORIGIN


  Query Match             80.0%;  Score 19.2;  DB 12;  Length 438;
  Best Local Similarity   62.5%;  Pred. No. 7e+02;
  Matches   15;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          |||: |:||: |||:||||::| |
Db     84 GCATACTACTACCATCCAATTCTA 61


RESULT   2
AV558212
LOCUS       AV558212                 379 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  AV558212 Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ092d03F 3', mRNA sequence.
ACCESSION   AV558212
VERSION     AV558212.1  GI:8729627
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQ092d03F"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /tissue_type="green siliques"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      104 a     95 c     62 g    118 t
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 9;  Length 379;
  Best Local Similarity   68.2%;  Pred. No. 9.8e+02;
  Matches   15;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucc 23
          ||:||: |:||||:|||| :||
Db    245 CATCCTTCTCCCATCCAAGTCC 266


RESULT   3
BH172929
LOCUS       BH172929                 434 bp    DNA     linear   GSS 03-OCT-2001
DEFINITION  SALK_006364 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_006364, DNA sequence.
ACCESSION   BH172929
VERSION     BH172929.1  GI:15911190
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Heller,C., Kim,C.J., Jeske,A., Koesema,E., Meyers,M.C., Parker
            ,H., Prednis,L., Shinn,P., Stevenson,D.K., Zimmerman,J. and Ecker
            ,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within 300 bases of the 3' end of
            At1g01920.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_006364"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT      135 a     86 c     66 g    147 t
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 12;  Length 434;
  Best Local Similarity   68.2%;  Pred. No. 1e+03;
  Matches   15;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucc 23
          ||:||: |:||||:|||| :||
Db    346 CATCCTTCTCCCATCCAAGTCC 367


RESULT   4
AG076330
LOCUS       AG076330                 661 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-070I01.R, genomic survey sequence.
ACCESSION   AG076330
VERSION     AG076330.1  GI:16628132
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-070I01.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 661)
```

```
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-070I01.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      140 a    194 c    129 g    197 t      1 others
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 12;  Length 661;
  Best Local Similarity   68.2%;  Pred. No. 1.1e+03;
  Matches   15;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 cauccuacucccauccaauucc 23
           || ||:||:||||:||| ::||
Db     300 CAACCTACTCCCATCCACTTCC 321


RESULT   5
AV329061
LOCUS       AV329061                286 bp    mRNA    linear   EST 11-NOV-1999
DEFINITION  AV329061 RIKEN full-length enriched, adult male medulla oblongata
            Mus musculus cDNA clone 6330500C11 3', mRNA sequence.
ACCESSION   AV329061
VERSION     AV329061.1  GI:6369113
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .286
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6330500C11"
                     /clone_lib="RIKEN full-length enriched, adult male medulla
                     oblongata"
                     /sex="male"
                     /tissue_type="medulla oblongata"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT       80 a    100 c     48 g     58 t
ORIGIN


  Query Match             76.7%;  Score 18.4;  DB 9;  Length 286;
  Best Local Similarity   65.0%;  Pred. No. 1.4e+03;
  Matches   13;  Conservative    6;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 auccuacucccauccaauuc 22
           |:||:||:||||:|| |::|
Db      48 ATCCTACTCCCATCCGATTC 67


RESULT   6
AI866251/c
LOCUS       AI866251                361 bp    mRNA    linear   EST 07-MAR-2000
DEFINITION  wl27f06.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2426147 3',
            mRNA sequence.
ACCESSION   AI866251
VERSION     AI866251.1  GI:5530347
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
```

```
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 2439   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2426147"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      102 a     65 c     90 g    104 t
ORIGIN


  Query Match             76.7%;  Score 18.4;  DB 9;  Length 361;
  Best Local Similarity   70.0%;  Pred. No. 1.4e+03;
  Matches   14;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 ccuacucccauccaaauucca 24
           ||:||:||||:||| ::|||
Db     216 CCTACTCCCATCCATTTCCA 197


RESULT   7
AQ248820
LOCUS       AQ248820                 202 bp    DNA     linear   GSS 06-OCT-1998
DEFINITION  T16N9-Sp6 TAMU Arabidopsis thaliana genomic clone T16N9, DNA
            sequence.
ACCESSION   AQ248820
VERSION     AQ248820.1  GI:3698903
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 202)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .202
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T16N9"
                     /clone_lib="TAMU"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                     ; Produced by Rod Wing"
BASE COUNT       64 a     51 c     37 g     47 t      3 others
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 12;  Length 202;
  Best Local Similarity   65.2%;  Pred. No. 1.5e+03;
  Matches   15;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcauccuacucccauccaaauucc 23
           ||| ||:||: | |:||||::||
Db      40 GCACCCTACTTCTATCCAATTCC 62


RESULT   8
BM089677
LOCUS       BM089677                 284 bp    mRNA    linear   EST 19-NOV-2001
DEFINITION  503497 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BM089677
VERSION     BM089677.1  GI:17000305
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 284)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 5  row: K  column: 6
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .284
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       46 a    104 c     94 g     40 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 10;  Length 284;
  Best Local Similarity   69.6%;  Pred. No. 1.6e+03;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy      1 gcauccuacucccauccaauucc 23
          |||:||  |:||||:|||| :||
Db    230 GCATCCAGCTCCCATCCAACTCC 252


RESULT  9
AA153510/c
LOCUS       AA153510                321 bp    mRNA    linear   EST 19-FEB-1997
DEFINITION  mq57b05.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:582801
            5', mRNA sequence.
ACCESSION   AA153510
VERSION     AA153510.1  GI:1725240
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 321)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:357449
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 315.
FEATURES             Location/Qualifiers
     source          1. .321
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:582801"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       67 a     89 c     88 g     77 t
ORIGIN


  Query Match           75.8%;  Score 18.2;  DB 9;  Length 321;
  Best Local Similarity   69.6%;  Pred. No. 1.7e+03;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          | :||:|| ||||:|||| :|||
Db    313 CTTCCTACCCCCATCCAACTCCA 291


RESULT  10
AZ442220
```

```
LOCUS       AZ442220                327 bp    DNA     linear   GSS 03-OCT-2000
DEFINITION  1M0234E19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0234E19 R, DNA sequence.
ACCESSION   AZ442220
VERSION     AZ442220.1  GI:10566233
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0234  row: E  column: 19
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 327.
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0234E19"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       62 a    108 c     90 g     67 t
ORIGIN


  Query Match           75.8%;  Score 18.2;  DB 12;  Length 327;
  Best Local Similarity   69.6%;  Pred. No. 1.7e+03;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          | :||:|| ||||:|||| :|||
Db    141 CTTCCTACCCCCATCCAACTCCA 163


RESULT  11
```

```
AW789229
LOCUS       AW789229                343 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  C00996-R Lambda Zap, Stratagene Blumeria graminis f. sp. hordei
            cDNA clone C00996, mRNA sequence.
ACCESSION   AW789229
VERSION     AW789229.1  GI:13900826
KEYWORDS    EST.
SOURCE      Blumeria graminis f. sp. hordei.
  ORGANISM  Blumeria graminis f. sp. hordei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
            Erysiphales; Erysiphaceae; Blumeria.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   Thomas,S.W., Rasmussen,S.W., Glaring,M.A., Rouster,J.A. and Oliver
            ,R.P.
  TITLE     Gene identification in the fungal pathogen Blumeria graminis by
            expressed sequence tag analysis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Rasmussen,S.W.
            Department of Yeast Genetics
            Carlsberg Laboratory
            10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark
            Tel: 45 3327 5230
            Fax: 45 3327 4766
            Email: swr@crc.dk
            High quality sequence stop: 343
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .343
                     /organism="Blumeria graminis f. sp. hordei"
                     /db_xref="taxon:62688"
                     /clone="C00996"
                     /clone_lib="Lambda Zap, Stratagene"
                     /cell_type="conidia"
                     /lab_host="Hordeum vulgare"
BASE COUNT       60 a     99 c     85 g     99 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 9;  Length 343;
  Best Local Similarity   65.2%;  Pred. No. 1.7e+03;
  Matches   15;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          ||:||:  :||||:||||: |||
Db     29 CATCCTTATCCCATCCAATGCCA 51


RESULT  12
BF601807
LOCUS       BF601807                411 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  266877 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF601807
VERSION     BF601807.1  GI:11699030
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 42  row: K  column: 6
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       64 a    131 c    133 g     83 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 10;  Length 411;
  Best Local Similarity   69.6%;  Pred. No. 1.7e+03;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucc 23
          |||:||  |:||||:|||| :||
Db    203 GCATCCAGCTCCCATCCAACTCC 225


RESULT  13
AI645377/c
LOCUS       AI645377                431 bp    mRNA    linear   EST 29-APR-1999
DEFINITION  mq57b05.y1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:582801
            5', mRNA sequence.
ACCESSION   AI645377
VERSION     AI645377.1  GI:4723852
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 418.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:582801"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```

```
                was primed with a Not I - oligo(dT) primer [5'
                TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTT
                3']; double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT7T3 vector. RNA
                provided by Dr. Bertrand Jordan. Library went through two
                rounds of normalization, and was constructed by Bento
                Soares and M.Fatima Bonaldo."
BASE COUNT       89 a    116 c    124 g    101 t      1 others
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 9;  Length 431;
  Best Local Similarity   69.6%;  Pred. No. 1.7e+03;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          | :||:|| ||||:|||| :|||
Db    347 CTTCCTACCCCCATCCAACTCCA 325


RESULT  14
BB860142/c
LOCUS       BB860142                462 bp    mRNA    linear   EST 26-NOV-2001
DEFINITION  BB860142 RIKEN full-length enriched, pooled cell lines Mus musculus
            cDNA clone G430012E17 5', mRNA sequence.
ACCESSION   BB860142
VERSION     BB860142.1  GI:17101596
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES            Location/Qualifiers
     source         1. .462
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /clone="G430012E17"
                    /clone_lib="RIKEN full-length enriched, pooled cell lines"
                    /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                    (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                    (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                    (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                    melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                    ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                    (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                    (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                    (tissue_type=bone marrow, cell_type=stroma cell,
                    cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                    cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                    cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                    cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                    cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                    ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                    BC3H1)"
BASE COUNT      109 a    121 c    161 g     71 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 9;  Length 462;
  Best Local Similarity   60.9%;  Pred. No. 1.8e+03;
  Matches   14;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          ||:| :  :||||:||||::|||
Db    294 CATCATTTTCCCATCCAATTCCA 272


RESULT  15
AQ978003/c
LOCUS       AQ978003                511 bp    DNA     linear   GSS 29-JAN-2000
DEFINITION  RPCI-23-339H5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-339H5,
            DNA sequence.
ACCESSION   AQ978003
VERSION     AQ978003.1  GI:6810304
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-339H5.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 339  row: H  column: 5
            Seq primer: SP6
            Class: BAC ends.
FEATURES            Location/Qualifiers.
     source         1. .511
```

```
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-339H5"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      147 a    114 c    125 g    125 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 12;  Length 511;
  Best Local Similarity   65.2%;  Pred. No. 1.8e+03;
  Matches   15;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 cauccuacucccauccaauucca 24
           || ||:| :||| :||||::|||
Db     275 CACCCTAATCCCTTCCAATTCCA 253


Search completed: June  2, 2002, 19:59:57
Job time: 14607 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 15:54:15 ; Search time 108.28 Seconds
(without alignments)
54.444 Million cell updates/sec

Title: US-09-663-600A-2_COPY_1_24
Perfect score: 24
Sequence: 1 gcauccuacucccauccaauucca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24	100.0	46	4	US-08-905-223-2	Sequence 2, Appli
	2	24	100.0	46	4	US-09-247-155-2	Sequence 2, Appli
	3	24	100.0	47	4	US-08-905-223-1	Sequence 1, Appli
	4	24	100.0	47	4	US-09-247-155-1	Sequence 1, Appli
	5	24	100.0	65	3	US-08-930-102A-1	Sequence 1, Appli
	6	23	95.8	46	3	US-08-930-102A-4	Sequence 4, Appli
	7	23	95.8	46	3	US-08-930-102A-5	Sequence 5, Appli
c	8	18.2	75.8	1312	2	US-08-841-349-6	Sequence 6, Appli
c	9	18.2	75.8	6960	2	US-08-841-349-3	Sequence 3, Appli
c	10	18.2	75.8	8176	2	US-08-841-349-5	Sequence 5, Appli
	11	16.2	67.5	250	3	US-08-301-162-13	Sequence 13, Appl
	12	16.2	67.5	250	4	US-09-461-240-13	Sequence 13, Appl
	13	16	66.7	305	1	US-08-170-294-5	Sequence 5, Appli
	14	16	66.7	305	2	US-08-664-855-5	Sequence 5, Appli
	15	16	66.7	305	3	US-09-049-289-5	Sequence 5, Appli
c	16	16	66.7	2358	1	US-08-339-152A-15	Sequence 15, Appl
c	17	16	66.7	2358	2	US-08-007-999B-2	Sequence 2, Appli
c	18	16	66.7	2358	2	US-08-689-276A-2	Sequence 2, Appli
	19	16	66.7	2493	3	US-08-945-056-3	Sequence 3, Appli
	20	16	66.7	3827	1	US-08-170-294-6	Sequence 6, Appli
	21	16	66.7	3827	2	US-08-664-855-6	Sequence 6, Appli
	22	16	66.7	3827	2	US-08-718-751-1	Sequence 1, Appli
	23	16	66.7	3827	3	US-09-049-289-6	Sequence 6, Appli
	24	16	66.7	6671	1	US-08-280-443-1	Sequence 1, Appli
	25	16	66.7	6671	1	US-08-457-459-1	Sequence 1, Appli
	26	16	66.7	6671	1	US-08-555-678-1	Sequence 1, Appli
	27	16	66.7	6671	5	PCT-US95-02275-1	Sequence 1, Appli
c	28	15.8	65.8	1664	1	US-07-863-169A-6	Sequence 6, Appli
c	29	15.8	65.8	1664	2	US-08-429-964-6	Sequence 6, Appli
c	30	15.8	65.8	1664	3	US-07-935-087-6	Sequence 6, Appli
c	31	15.8	65.8	1664	5	PCT-US93-08062-6	Sequence 6, Appli
	32	15.6	65.0	567	4	US-09-385-982-427	Sequence 427, App
	33	15.6	65.0	568	4	US-09-247-155-136	Sequence 136, App
	34	15.6	65.0	569	3	US-09-109-204-15	Sequence 15, Appl
c	35	15.6	65.0	1843	3	US-08-718-738-1	Sequence 1, Appli
c	36	15.6	65.0	1843	4	US-09-221-844-1	Sequence 1, Appli
c	37	15.6	65.0	1843	5	PCT-US95-03323A-1	Sequence 1, Appli
	38	15.6	65.0	2294	3	US-08-964-700A-1	Sequence 1, Appli
	39	15.6	65.0	2411	3	US-09-188-930-75	Sequence 75, Appl
	40	15.6	65.0	2411	3	US-09-188-930-256	Sequence 256, App
	41	15.2	63.3	1288	4	US-09-296-754-1	Sequence 1, Appli
	42	15	62.5	50	4	US-09-315-886C-15	Sequence 15, Appl
c	43	15	62.5	87	3	US-08-816-346-43	Sequence 43, Appl
c	44	15	62.5	87	3	US-09-335-411-43	Sequence 43, Appl
c	45	15	62.5	153	3	US-08-816-346-39	Sequence 39, Appl

ALIGNMENTS

```
RESULT 1
US-08-905-223-2
; Sequence 2, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste D.
;    APPLICANT:  Duelert, Aymeric
;    APPLICANT:  Lacroix, Bruno
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,223
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  46 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  Other nucleic acid
US-08-905-223-2


  Query Match           100.0%;  Score 24;  DB 4;  Length 46;
  Best Local Similarity  100.0%;  Pred. No. 0.004;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcauccuacucccauccaauucca 24
           ||||||||||||||||||||||||
Db       1 GCAUCCUACUCCCAUCCAAUUCCA 24
```

```
RESULT   2
US-09-247-155-2
; Sequence 2, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 2
;   LENGTH: 46
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: in vitro transcription product
US-09-247-155-2


  Query Match             100.0%;  Score 24;  DB 4;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      1 gcauccuacucccauccaauucca 24


RESULT   3
US-08-905-223-1
; Sequence 1, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste D.
;    APPLICANT:  Duelert, Aymeric
;    APPLICANT:  Lacroix, Bruno
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,223
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  47 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  Other nucleic acid
;    FEATURE:
;      NAME/KEY:  Cap
;      LOCATION:  1
;      OTHER INFORMATION:  m7Gppp added to 1
US-08-905-223-1


  Query Match             100.0%;  Score 24;  DB 4;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      2 GCAUCCUACUCCCAUCCAAUUCCA 25


RESULT   4
US-09-247-155-1
; Sequence 1, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 47
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: in vitro transcription product
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: 1
;   OTHER INFORMATION: m7g
US-09-247-155-1


  Query Match             100.0%;  Score 24;  DB 4;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      2 gcauccuacucccauccaauucca 25


RESULT   5
US-08-930-102A-1
; Sequence 1, Application US/08930102A
; Patent No. 6022715
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Dumas, Jean-Baptiste Milne Edwards
;    APPLICANT:  Merenkova, Irena Nicolaevna
;    TITLE OF INVENTION:  METHOD FOR THE SPECIFIC COUPLING OF THE CAP
;    TITLE OF INVENTION:  OF THE 5'END OF AN mRNA FRAGMENT AND PREPARTION OF mRNA AND OF
;    TITLE OF INVENTION:  COMPLETE cDNA
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, 16th Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/930,102A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR96/00651
;      FILING DATE:  29-APR-1996
;      APPLICATION NUMBER:  FR95/05221
;      FILING DATE:  02-MAY-1995
;      APPLICATION NUMBER:  FR95/09467
;      FILING DATE:  03-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ned A. Israelsen
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.017APC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  65 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-930-102A-1


  Query Match             100.0%;  Score 24;  DB 3;  Length 65;
  Best Local Similarity   75.0%;  Pred. No. 0.0043;
  Matches   18;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcauccuacucccauccaauucca 24
           |||:||:||:||||:||||::|||
Db      20 GCATCCTACTCCCATCCAATTCCA 43


RESULT   6
US-08-930-102A-4
; Sequence 4, Application US/08930102A
; Patent No. 6022715
;  GENERAL INFORMATION:
;    APPLICANT:  Dumas, Jean-Baptiste Milne Edwards
;    APPLICANT:  Merenkova, Irena Nicolaevna
;    TITLE OF INVENTION:  METHOD FOR THE SPECIFIC COUPLING OF THE CAP
;    TITLE OF INVENTION:  OF THE 5'END OF AN mRNA FRAGMENT AND PREPARTION OF mRNA AND OF
;    TITLE OF INVENTION:  COMPLETE cDNA
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, 16th Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/930,102A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR96/00651
;      FILING DATE:  29-APR-1996
;      APPLICATION NUMBER:  FR95/05221
;      FILING DATE:  02-MAY-1995
;      APPLICATION NUMBER:  FR95/09467
;      FILING DATE:  03-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ned A. Israelsen
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.017APC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  46 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  RNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1
;      OTHER INFORMATION:  /label= m7Gppp
;      OTHER INFORMATION:  /note= "N is m7GpppG"
US-08-930-102A-4


  Query Match             95.8%;  Score 23;  DB 3;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cauccuacucccauccaauucca 24
           |||||||||||||||||||||||
Db       2 CAUCCUACUCCCAUCCAAUUCCA 24


RESULT   7
US-08-930-102A-5
; Sequence 5, Application US/08930102A
; Patent No. 6022715
;  GENERAL INFORMATION:
;    APPLICANT:  Dumas, Jean-Baptiste Milne Edwards
;    APPLICANT:  Merenkova, Irena Nicolaevna
;    TITLE OF INVENTION:  METHOD FOR THE SPECIFIC COUPLING OF THE CAP
;    TITLE OF INVENTION:  OF THE 5'END OF AN mRNA FRAGMENT AND PREPARTION OF mRNA AND
;    TITLE OF INVENTION:  COMPLETE cDNA
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, 16th Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/930,102A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR96/00651
;      FILING DATE:  29-APR-1996
;      APPLICATION NUMBER:  FR95/05221
;      FILING DATE:  02-MAY-1995
;      APPLICATION NUMBER:  FR95/09467
;      FILING DATE:  03-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ned A. Israelsen
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.017APC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  46 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  RNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1
;      OTHER INFORMATION:  /label= ppp
;      OTHER INFORMATION:  /note= "N is pppG"
US-08-930-102A-5


  Query Match             95.8%;  Score 23;  DB 3;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          |||||||||||||||||||||||
Db      2 CAUCCUACUCCCAUCCAAUUCCA 24


RESULT   8
US-08-841-349-6/c
; Sequence 6, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
;  APPLICANT: MISHRA, LOPA
;  TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT. . .
;  FILE REFERENCE: XX/PO4470US0
;  CURRENT APPLICATION NUMBER: US/08/841,349B
;  CURRENT FILING DATE: 1997-04-30
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 1312
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (402)..(1061)
US-08-841-349-6


  Query Match             75.8%;  Score 18.2;  DB 2;  Length 1312;
  Best Local Similarity   69.6%;  Pred. No. 4.3;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          | :||:|| ||||:|||| :|||
Db    347 CTTCCTACCCCCATCCAACTCCA 325


RESULT   9
US-08-841-349-3/c
; Sequence 3, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
;  APPLICANT: MISHRA, LOPA
;  TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT. . .
;  FILE REFERENCE: XX/PO4470US0
;  CURRENT APPLICATION NUMBER: US/08/841,349B
;  CURRENT FILING DATE: 1997-04-30
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 6960
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (333)..(6794)
US-08-841-349-3


  Query Match             75.8%;  Score 18.2;  DB 2;  Length 6960;
  Best Local Similarity   69.6%;  Pred. No. 5.7;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          | :||:|| ||||:|||| :|||
Db    278 CTTCCTACCCCCATCCAACTCCA 256


RESULT  10
US-08-841-349-5/c
; Sequence 5, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
;  APPLICANT: MISHRA, LOPA
;  TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT. . .
;  FILE REFERENCE: XX/PO4470US0
;  CURRENT APPLICATION NUMBER: US/08/841,349B
;  CURRENT FILING DATE: 1997-04-30
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 8176
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-08-841-349-5


  Query Match             75.8%;  Score 18.2;  DB 2;  Length 8176;
  Best Local Similarity   69.6%;  Pred. No. 5.8;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          | :||:|| ||||:|||| :|||
Db    278 CTTCCTACCCCCATCCAACTCCA 256


RESULT  11
US-08-301-162-13
; Sequence 13, Application US/08301162
; Patent No. 6022546
;  GENERAL INFORMATION:
;    APPLICANT:  Knapp, Stefan
;    APPLICANT:  Ziegelmaier, Robert
;    APPLICANT:  Kupper, Hans
;    TITLE OF INVENTION:  Toxoplasma Gondii Antigens, The
;    TITLE OF INVENTION:  Preparation Thereof and the Use Thereof
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
```

```
;       ADDRESSEE:  Dunner
;       STREET:  1300 I Street, N.W., Suite 700
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20005-3315
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/301,162
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/167,128
;       FILING DATE:
;       APPLICATION NUMBER:  US 07/623,086
;       FILING DATE:  06-DEC-1990
;       APPLICATION NUMBER:  DE P3940598.2
;       FILING DATE:  08-DEC-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fleshner, Raz E.
;       REGISTRATION NUMBER:  34,331
;       REFERENCE/DOCKET NUMBER:  02481.1005-00000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-408-4000
;       TELEFAX:  202-408-4400
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  250 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-301-162-13


  Query Match             67.5%;  Score 16.2;  DB 3;  Length 250;
  Best Local Similarity   57.1%;  Pred. No. 29;
  Matches   12;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauuc 22
          | :||: |:||||:|| |::|
Db    178 CTTCCTTCTCCCATCCGATTC 198


RESULT  12
US-09-461-240-13
; Sequence 13, Application US/09461240
; Patent No. 6326008
;   GENERAL INFORMATION:
;       APPLICANT: Knapp, Stefan
;                  Ziegelmaier, Robert
;                  Kupper, Hans
;       TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
;                           Preparation Thereof and the Use Thereof
;       NUMBER OF SEQUENCES: 20
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;                       Dunner
;            STREET: 1300 I Street, N.W., Suite 700
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: USA
;            ZIP: 20005-3315
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/461,240
;            FILING DATE: 16-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/301,162
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: US/08/167,128
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: US 07/623,086
;            FILING DATE: 06-DEC-1990
;            APPLICATION NUMBER: DE P3940598.2
;            FILING DATE: 08-DEC-1989
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Fleshner, Raz E.
;            REGISTRATION NUMBER: 34,331
;            REFERENCE/DOCKET NUMBER: 02481.1005-00000
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 202-408-4000
;            TELEFAX: 202-408-4400
;   INFORMATION FOR SEQ ID NO: 13:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 250 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-461-240-13


  Query Match             67.5%;  Score 16.2;  DB 4;  Length 250;
  Best Local Similarity   57.1%;  Pred. No. 29;
  Matches   12;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauuc 22
          | :||: |:||||:|| |::|
Db    178 CTTCCTTCTCCCATCCGATTC 198


RESULT  13
US-08-170-294-5
; Sequence 5, Application US/08170294
; Patent No. 5589614
;  GENERAL INFORMATION:
;    APPLICANT:  BRIDGES, IAN G.
;    APPLICANT:  BRIGHT, SIMON W.J.
;    APPLICANT:  GREENLAND, ANDREW J.
;    APPLICANT:  HOLT, DAVID C.
;    APPLICANT:  JEPSON, IAN
;    APPLICANT:  SCHUCH, WOLFGANG W.
;    TITLE OF INVENTION:  PLANT-DERIVED ENZYME AND DNA SEQUENCES
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/170,294
;      FILING DATE:  30-DEC-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/GB92/01187
;      FILING DATE:  01-JUL-1992
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  GB 9114259.6
;      FILING DATE:  02-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  204218/SEE36438/UST
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  PRIMER EXTENSION MAP FIGURE 5
US-08-170-294-5


  Query Match             66.7%;  Score 16;  DB 1;  Length 305;
  Best Local Similarity   66.7%;  Pred. No. 38;
  Matches   16;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||  |  || ||||:||||::|||
Db    167 GCTGCACACACCCATCCAATTCCA 190


RESULT  14
US-08-664-855-5
; Sequence 5, Application US/08664855
; Patent No. 5866792
;  GENERAL INFORMATION:
;    APPLICANT:  BRIDGES, IAN G
;    APPLICANT:  BRIGHT, SIMON WJ
;    APPLICANT:  GREENLAND, ANDREW J
;    APPLICANT:  HOLT, DAVID C
;    APPLICANT:  JEPSON, IAN
;    APPLICANT:  SCHUCH, WOLFGANG W
;    TITLE OF INVENTION:  PLANT-DERIVED ENZYME AND DNA SEQUENCES,
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/664,855
;      FILING DATE:  17-JUN-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/170,294
;      FILING DATE:  30-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9114259.6
;      FILING DATE:  02-JUL-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB92/01187
;      FILING DATE:  01-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  224452/SEE36438USTD1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  PRIMER EXTENSION MAP FIGURE 5
US-08-664-855-5


  Query Match             66.7%;  Score 16;  DB 2;  Length 305;
  Best Local Similarity   66.7%;  Pred. No. 38;
  Matches   16;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||  |  || ||||:||||::|||
Db    167 GCTGCACACACCCATCCAATTCCA 190


RESULT  15
US-09-049-289-5
; Sequence 5, Application US/09049289
; Patent No. 6066456
;  GENERAL INFORMATION:
;    APPLICANT:  BRIDGES, IAN G.
;    APPLICANT:  BRIGHT, SIMON W.J.
;    APPLICANT:  GREENLAND, ANDREW J.
;    APPLICANT:  HOLT, DAVID C.
;    APPLICANT:  JEPSON, IAN
;    APPLICANT:  SCHUCH, WOLFGANG W.
;    TITLE OF INVENTION:  PLANT-DERIVED ENZYME AND DNA SEQUENCES
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/049,289
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/170,294
;      FILING DATE:  30-DEC-1993
;      APPLICATION NUMBER:  WO PCT/GB92/01187
;      FILING DATE:  01-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9114259.6
;      FILING DATE:  02-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  204218/SEE36438/UST
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  PRIMER EXTENSION MAP FIGURE 5
US-09-049-289-5


  Query Match             66.7%;  Score 16;  DB 3;  Length 305;
  Best Local Similarity   66.7%;  Pred. No. 38;
  Matches   16;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcauccuacucccauccaauucca 24
           ||  |  || ||||:||||::|||
Db     167 GCTGCACACACCCATCCAATTCCA 190


Search completed: June  2, 2002, 18:58:39
Job time: 11064 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 16:03:15 ; Search time 523.16 Seconds
(without alignments)
78.764 Million cell updates/sec

Title: US-09-663-600A-2_COPY_1_24
Perfect score: 24
Sequence: 1 gcauccuacucccauccaauucca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24	100.0	46	20	AAZ40750	Oligonucleotide -C
2	24	100.0	46	20	AAX88171	Oligoribonucleotid
3	24	100.0	46	20	AAX97544	Oligonucleotide Ca
4	24	100.0	46	20	AAX19963	Oligoribonucleotid
5	24	100.0	46	20	AAX51429	Oligonucleotide -C
6	24	100.0	46	20	AAX39410	Human secreted pro
7	24	100.0	46	20	AAX40408	Oligoribonucleotid
8	24	100.0	46	20	AAX51757	Uncapped mRNA for
9	24	100.0	46	20	AAX26652	Oligoribonucleotid
10	24	100.0	46	20	AAX41349	Oligoribonucleotid
11	24	100.0	46	20	AAX30063	Oligoribonucleotid
12	24	100.0	47	17	AAT43581	5'-capped mRNA for
13	24	100.0	47	20	AAZ40749	Oligonucleotide +C
14	24	100.0	47	20	AAX88170	Oligoribonucleotid
15	24	100.0	47	20	AAX97543	Oligonucleotide Ca
16	24	100.0	47	20	AAX19962	Oligoribonucleotid
17	24	100.0	47	20	AAX51428	Oligonucleotide +C
18	24	100.0	47	20	AAX39409	Human secreted pro
19	24	100.0	47	20	AAX40407	Oligoribonucleotid
20	24	100.0	47	20	AAX51756	Capped mRNA for ES
21	24	100.0	47	20	AAX26651	Oligoribonucleotid
22	24	100.0	47	20	AAX41348	Oligoribonucleotid
23	24	100.0	47	20	AAX30062	Oligoribonucleotid
c 24	18.2	75.8	1312	20	AAV64413	Mouse elf-3 cDNA.
c 25	18.2	75.8	6991	20	AAV64412	Mouse elf-2 cDNA.
c 26	18.2	75.8	8176	20	AAV64411	Mouse elf-1 cDNA.
c 27	17.8	74.2	1161	21	AAC49972	Arabidopsis thalia
c 28	17.8	74.2	1542	21	AAC41500	Arabidopsis thalia
c 29	17.8	74.2	1605	21	AAC43079	Arabidopsis thalia
c 30	17.8	74.2	1650	21	AAC49646	Arabidopsis thalia
31	17.6	73.3	417	21	AAC06005	Human secreted pro
32	17.6	73.3	3255	23	AAS68839	DNA encoding novel
c 33	17.2	71.7	633	22	AAH05985	Human cDNA clone (
c 34	17.2	71.7	879	22	AAH65420	C glutamicum codin
c 35	17.2	71.7	2119	22	AAH17697	Human cDNA sequenc
c 36	17.2	71.7	5672	22	AAH15193	Human cDNA sequenc
c 37	17.2	71.7	6802	22	AAS46282	Tumour suppressor
c 38	17.2	71.7	6802	24	AAS61068	Human gene regulat
c 39	17.2	71.7	349980	22	AAH68525	C glutamicum codin
40	16.8	70.0	322	20	AAV86922	EST clone BK280.
c 41	16.8	70.0	329	21	AAC26424	Human secreted pro
42	16.8	70.0	3797	23	ABL21119	Drosophila melanog
43	16.8	70.0	3858	23	ABL11191	Drosophila melanog
c 44	16.8	70.0	4087	22	AAS27712	DNA encoding novel
45	16.8	70.0	6461	23	ABL11190	Drosophila melanog

ALIGNMENTS

RESULT 1
AAZ40750
ID AAZ40750 standard; RNA; 46 BP.
XX
AC AAZ40750;
XX
DT 18-JAN-2000 (first entry)
XX
DE Oligonucleotide -Cap for Secreted protein EST isolation.
XX
KW PCR primer; secreted protein; fingerprint identification technique;
KW chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW hypertension; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9940189-A2.
XX
PD 12-AUG-1999.
XX
PF 09-FEB-1999; 99WO-IB00282.
XX
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
PR 10-AUG-1998; 98US-0096116.
PR 04-SEP-1998; 98US-0099273.
XX
PA (GEST) GENSET.

XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-600966/51.
XX
PT Extended cDNAs useful for expressing secreted proteins and to obtain
PT specific antibodies -
XX
PS Example 2; Page 12; 244pp; English.
XX
CC This sequence represents a PCR primer used within the course of the
CC invention. The invention relates to 70 nucleic acids encoding human
CC secreted proteins. The extended cDNAs (or genomic DNAs obtainable from
CC them) may be used to prepare PCR primers and probes. These are useful for
CC forensic matching or positive identification by DNA sequencing. They may
CC also be used in alternative fingerprint identification techniques.
CC Antibodies against the proteins encoded by the extended cDNAs are useful
CC in identification of tissue types or cell species, as well as identifying
CC tissue specific soluble proteins. The sequences can be used for
CC chromosome mapping and identification of genes associated with hereditary
CC diseases or drug response. signal sequences from the cDNAs can be used in
CC construction of secretion vectors. Other sequences derived from the
CC extended cDNAs can be used to clone upstream genomic DNA sequences
CC including promoters. This is in turn useful for identifying proteins that
CC interact with promoter sequences. Some of the proteins may be useful in
CC diagnosing and treating several disorders including, but not limited to:
CC cancer, hyperlipidaemia, cardiovascular and neurodegenerative disorders,
CC autoimmune diseases, and rheumatic diseases, embryogenic disorders,
CC hypertension, renal injury, amino acidurias, hypoglycaemia, male rat
CC infertility and myopathies.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauccaauucca 24
||||||||||||||||||||||||
Db 1 gcauccuacucccauccaauucca 24

RESULT 2
AAX88171
ID AAX88171 standard; RNA; 46 BP.
XX
AC AAX88171;
XX
DT 23-SEP-1999 (first entry)
XX
DE Oligoribonucleotide Cap5'-ppp.
XX
KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW diagnostic; gene therapy; chromosome mapping; secretion vector; primer;
KW ss.
XX
OS Synthetic.
XX
PN WO9925825-A2.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-IB01862.
XX
PR 04-SEP-1998; 98US-0099273.
PR 13-NOV-1997; 97US-0066677.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
PR 10-AUG-1998; 98US-0096116.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-347472/29.
XX
PT Extended cDNAs encoding secreted proteins
XX
PS Example 2; Page 131; 307pp; English.
XX
CC This invention describes novel nucleic acid sequences of extended cDNAs
CC (see AAX97813-X97906) which encode human secreted proteins (see
CC AAY36129-Y36222) and which have cytostatic, thrombotic and osteopathic
CC activity. The extended cDNAs can be used to express secreted proteins
CC or parts of them or to obtain antibodies capable of binding to the
CC secreted proteins. They may also be used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. Uses also include design
CC of expression vectors and secretion vectors. This sequence represents
CC an oligoribonucleotide primer used in the method of the invention.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauccaauucca 24
||||||||||||||||||||||||
Db 1 gcauccuacucccauccaauucca 24

RESULT 3
AAX97544
ID AAX97544 standard; RNA; 46 BP.
XX
AC AAX97544;
XX
DT 13-SEP-1999 (first entry)
XX
DE Oligonucleotide Cap- for secreted protein coding sequence isolation.
XX
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9931236-A2.
XX
PD 24-JUN-1999.
XX
PF 17-DEC-1998; 98WO-IB02122.
XX
PR 10-AUG-1998; 98US-0096116.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-385906/32.
XX
PT New isolated human secreted proteins
XX
PS Example 2; Page 12; 516pp; English.
XX
CC This sequence represents an oligonucleotide used to isolate the extended
CC human secreted protein coding sequences of the invention. The secreted

CC proteins can be used in treating or controlling a variety of human
CC conditions. The secreted proteins may act as cytokines or may affect
CC cellular proliferation or differentiation or may act as immune system
CC regulators, haematopoiesis regulators, tissue growth regulators,
CC regulators of reproductive hormones or cell movement or have
CC chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or tumour
CC inhibition activity. The DNAs can be used in forensic procedures to
CC identify individuals or in diagnostic procedures to identify individuals
CC having genetic diseases resulting from abnormal expression of the genes
CC corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauccaauucca 24
||||||||||||||||||||||||
Db 1 gcauccuacucccauccaauucca 24

RESULT 4
AAX19963
ID AAX19963 standard; RNA; 46 BP.
XX
AC AAX19963;
XX
DT 16-JUN-1999 (first entry)
XX
DE Oligoribonucleotide -Cap SEQ ID NO:2.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ss.
XX
OS Synthetic.
XX
PN WO9906439-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01233.
XX
PR 01-AUG-1997; 97US-0904468.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153700/13.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from liver, lung, large intestine, colon,
PT thyroid and pancreas tissue
XX
PS Example 2; Page 15; 398pp; English.
XX
CC AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY11533 to
CC AAY11679, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell. The
CC present sequence represents an oligoribonucleotide used in an example
CC from the present invention.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauccaauucca 24
||||||||||||||||||||||||
Db 1 gcauccuacucccauccaauucca 24

RESULT 5
AAX51429
ID AAX51429 standard; RNA; 46 BP.
XX
AC AAX51429;
XX
DT 21-JUN-1999 (first entry)
XX
DE Oligonucleotide -Cap.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ss.
XX
OS Synthetic.
XX
PN WO9906549-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01231.
XX
PR 01-AUG-1997; 97US-0905279.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153779/13.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS Example 2; Page 144; 522pp; English.
XX
CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12681 to
CC AAY12913, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products

CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell. This
CC sequence represents an oligonucleotide -Cap, used in the isolation of the
CC 5' EST sequences of the invention.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauccaauucca 24
||||||||||||||||||||||||
Db 1 gcauccuacucccauccaauucca 24

RESULT 6
AAX39410
ID AAX39410 standard; RNA; 46 BP.
XX
AC AAX39410;
XX
DT 21-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST detecting oligoribonucleotide #2.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ss.
XX
OS Synthetic.
XX
PN WO9906551-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01235.
XX
PR 01-AUG-1997; 97US-0905133.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153781/13.
XX
PT New nucleic acids encoding human secreted - proteins obtained from
PT cDNA libraries prepared from substantia nigra, cerebellum, surrenals
PT and fetal brain tissue
XX
PS Example 2; Page 15; 434pp; English.
XX
CC This invention describes 5' expressed sequence tags (ESTs) represented
CC in AAX39440 to AAX39597, which encode the human secreted proteins
CC represented in AAY11374-Y11531. The proteins given represent the signal
CC peptide and an N-terminal fragment of a secreted protein. The nucleic
CC acid sequences can be used for producing secreted human gene products.
CC They can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell. This
CC sequence represents an oligoribonucleotide which is used in the method of
CC the invention.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauccaauucca 24
||||||||||||||||||||||||
Db 1 gcauccuacucccauccaauucca 24

RESULT 7
AAX40408
ID AAX40408 standard; RNA; 46 BP.
XX
AC AAX40408;
XX
DT 18-JUN-1999 (first entry)
XX
DE Oligoribonucleotide -Cap.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9906550-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01232.
XX
PR 01-AUG-1997; 97US-0905144.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153780/13.
XX
PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
PS Example 2; Page 15; 675pp; English.
XX
CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins expressed in prostate, and encode the proteins
CC given in AAY11716 to AAY11993 respectively. The proteins given represent
CC the signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation and differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptides can be used for

CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauccaauucca 24
||||||||||||||||||||||||
Db 1 gcauccuacucccauccaauucca 24

RESULT 8
AAX51757
ID AAX51757 standard; RNA; 46 BP.
XX
AC AAX51757;
XX
DT 22-JUN-1999 (first entry)
XX
DE Uncapped mRNA for EST sequence of human secreted protein.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ss.
XX
OS Homo sapiens.
XX
PN WO9906552-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01236.
XX
PR 01-AUG-1997; 97US-0905223.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153782/13.
XX
PT New isolated brain-derived nucleic acids - used to develop products
PT which may have cytokine, immune, regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
PS Example 2; Page 15; 577pp; English.
XX
CC AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12987 to
CC AAY13219, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
CC This sequence was used in a method to isolate the 5' ESTs of the genes
CC encoding the human secreted proteins of the invention.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauccaauucca 24
||||||||||||||||||||||||
Db 1 gcauccuacucccauccaauucca 24

RESULT 9
AAX26652
ID AAX26652 standard; RNA; 46 BP.
XX
AC AAX26652;
XX
DT 18-JUN-1999 (first entry)
XX
DE Oligoribonucleotide used to identify 5' EST sequences.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ss.
XX
OS Synthetic.
XX
PN WO9906554-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01238.
XX
PR 01-AUG-1997; 97US-0905134.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153784/13.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
XX
PS Example 2; Page 15; 622pp; English.
XX
CC The specification describes 5' expressed sequence tags (ESTs, see
CC AAX40826-X41093) for human secreted proteins (see AAY01602 and AAY11994-
CC AAY12260). The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell. The present sequence is used
CC in the course of the invention.
XX
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

```
  Query Match           100.0%;  Score 24;  DB 20;  Length 46;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      1 gcauccuacucccauccaauucca 24


RESULT  10
AAX41349
ID   AAX41349 standard; RNA; 46 BP.
XX
AC   AAX41349;
XX
DT   22-JUN-1999  (first entry)
XX
DE   Oligoribonucleotide, SEQ ID NO: 2 from WO9906553.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; antiinflammatory; tumour inhibition; antitumour; ss.
XX
OS   Homo sapiens.
XX
PN   WO9906553-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01237.
XX
PR   01-AUG-1997;   97US-0905051.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153783/13.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from umbilical cord, lymph ganglia,
PT   lymphocytes and placental tissue
XX
PS   Example 2; Page 15; 411pp; English.
XX
CC   The patent relates to sequences of 5' ESTs derived from mRNAs
CC   encoding secreted proteins. The nucleic acid sequences can
CC   be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, antiinflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
XX
SQ   Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;


  Query Match           100.0%;  Score 24;  DB 20;  Length 46;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      1 gcauccuacucccauccaauucca 24


RESULT  11
AAX30063
ID   AAX30063 standard; RNA; 46 BP.
XX
AC   AAX30063;
XX
DT   17-JUN-1999  (first entry)
XX
DE   Oligoribonucleotide -Cap SEQ ID NO:2.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ss.
XX
OS   Synthetic.
XX
PN   WO9906548-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01222.
XX
PR   01-AUG-1997;   97US-0905135.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153778/13.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
XX
PS   Example 2; Page 15; 824pp; English.
XX
CC   AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12261 to
CC   AAY12514, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell. The
CC   present sequence represents an oligoribonucleotide used in an example
CC   from the present invention.
XX
SQ   Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;


  Query Match           100.0%;  Score 24;  DB 20;  Length 46;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      1 gcauccuacucccauccaauucca 24
```

```
RESULT  12
AAT43581
ID   AAT43581 standard; mRNA; 47 BP.
XX
AC   AAT43581;
XX
DT   04-AUG-1997  (first entry)
XX
DE   5'-capped mRNA for coupling to biotin label via amino linker.
XX
KW   Messenger RNA; guanosine 5'-cap; label; immobilisation; capture;
KW   polymerase chain reaction; transcription template; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag=  a
FT                   /mod_base= m7g
FT                   /function= cap
FT                   /note= "linked to adjacent nucleotide via
FT                          (5')ppp(5') linkage"
XX
PN   WO9634981-A2.
XX
PD   07-NOV-1996.
XX
PF   29-APR-1996;   96WO-FR00651.
XX
PR   03-AUG-1995;   95FR-0009467.
PR   02-MAY-1995;   95FR-0005221.
XX
PA   (GENQ ) GENSET.
XX
PI   Dumas Milne Edwards JG,  Nicolaevna Merenkova I;
XX
DR   WPI; 1996-506181/50.
XX
PT   Specific coupling of the 5' cap of mRNA to amino-functionalised cpd.
PT   - by eliminating 3' diol, oxidn. of cap diol to aldehyde and
PT   reaction with amine, e.g. for isolation of complete RNA, labelling
PT   etc.
XX
PS   Example 1; Page 20; 49pp; French.
XX
CC   The first step in a new method for specifically coupling the cap of
CC   the 5'-end of a eukaryotic mRNA to an amino-functionalised compound
CC   involves specifically modifying the 3'-end of the mRNA so that the
CC   last base no longer contains OH groups at the 2' and 3' positions.
CC   Then, the 2', 3'-cis diol of the methyl guanosine residue at the 5'-end
CC   can be oxidised to form a 2',3'-dialdehyde which is ultimately coupled
CC   with the amino group of the amino-functionalised compound. The method
CC   is used to label specifically at the 5'-cap, to isolate the 5'-end
CC   of mRNA in a sample, to produce the 3'-end of cDNA, to produce double
CC   stranded cDNA complementary to the 5'-end of mRNA or to capture mRNA-
CC   binding proteins. In a specific example of the coupling method, a
CC   double-stranded template was prepared by PCR amplification using a
CC   5'-primer containing the T7 RNA polymerase promoter (AAT43579) and
CC   a 3'-primer (AAT43580). When the template was transcribed in the
CC   presence of cap analogue m7G(5')ppp(5')G, a capped RNA transcript
CC   having the present sequence was produced. The 5'-capped mRNA was
CC   coupled to a biotin label via a hydrazine linker.
XX
SQ   Sequence 47 BP; 10 A; 24 C; 2 G; 11 U; 0 other;
```

```
  Query Match           100.0%;  Score 24;  DB 17;  Length 47;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      2 gcauccuacucccauccaauucca 25
```

```
RESULT  13
AAZ40749
ID   AAZ40749 standard; RNA; 47 BP.
XX
AC   AAZ40749;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Oligonucleotide +Cap for Secreted protein EST isolation.
XX
KW   PCR primer; secreted protein; fingerprint identification technique;
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-600966/51.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Example 2; Page 12; 244pp; English.
XX
CC   This sequence represents a PCR primer used within the course of the
CC   invention. The invention relates to 70 nucleic acids encoding human
CC   secreted proteins. The extended cDNAs (or genomic DNAs obtainable from
CC   them) may be used to prepare PCR primers and probes. These are useful for
CC   forensic matching or positive identification by DNA sequencing. They may
CC   also be used in alternative fingerprint identification techniques.
CC   Antibodies against the proteins encoded by the extended cDNAs are useful
CC   in identification of tissue types or cell species, as well as identifying
CC   tissue specific soluble proteins. The sequences can be used for
CC   chromosome mapping and identification of genes associated with hereditary
CC   diseases or drug response. signal sequences from the cDNAs can be used in
CC   construction of secretion vectors. Other sequences derived from the
CC   extended cDNAs can be used to clone upstream genomic DNA sequences
CC   including promoters. This is in turn useful for identifying proteins that
CC   interact with promoter sequences. Some of the proteins may be useful in
CC   diagnosing and treating several disorders including, but not limited to:
CC   cancer, hyperlipidaemia, cardiovascular and neurodegenerative disorders,
CC   autoimmune diseases, and rheumatic diseases, embryogenic disorders,
CC   hypertension, renal injury, amino acidurias, hypoglycaemia, male rat-
CC   infertility and myopathies.
XX
SQ   Sequence 47 BP; 10 A; 24 C; 1 G; 11 U; 1 other;
```

```
  Query Match           100.0%;  Score 24;  DB 20;  Length 47;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      2 gcauccuacucccauccaauucca 25


RESULT 14
AAX88170
ID   AAX88170 standard; RNA; 47 BP.
XX
AC   AAX88170;
XX
DT   23-SEP-1999  (first entry)
XX
DE   Oligoribonucleotide 5'm7Gppp.
XX
KW   Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW   diagnostic; gene therapy; chromosome mapping; secretion vector; primer;
KW   ss.
XX
OS   Synthetic.
XX
PN   WO9925825-A2.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-IB01862.
XX
PR   04-SEP-1998;   98US-0099273.
PR   13-NOV-1997;   97US-0066677.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-347472/29.
XX
PT   Extended cDNAs encoding secreted proteins
XX
PS   Example 2; Page 131; 307pp; English.
XX
CC   This invention describes novel nucleic acid sequences of extended cDNAs
CC   (see AAX97813-X97906) which encode human secreted proteins (see
CC   AAY36129-Y36222) and which have cytostatic, thrombotic and osteopathic
CC   activity. The extended cDNAs can be used to express secreted proteins
CC   or parts of them or to obtain antibodies capable of binding to the
CC   secreted proteins. They may also be used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. Uses also include design
CC   of expression vectors and secretion vectors. This sequence represents
CC   an oligoribonucleotide primer used in the method of the invention.
XX
SQ   Sequence 47 BP; 10 A; 24 C; 2 G; 11 U; 0 other;

  Query Match           100.0%;  Score 24;  DB 20;  Length 47;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      2 gcauccuacucccauccaauucca 25


RESULT 15
AAX97543
ID   AAX97543 standard; RNA; 47 BP.
XX
AC   AAX97543;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Oligonucleotide Cap+ for secreted protein coding sequence isolation.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag=  a
FT                   /mod_base= 7-methylguanosine
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
XX
PT   New isolated human secreted proteins
XX
PS   Example 2; Page 12; 516pp; English.
XX
CC   This sequence represents an oligonucleotide used to isolate the extended
CC   human secreted protein coding sequences of the invention. The secreted
CC   proteins can be used in treating or controlling a variety of human
CC   conditions. The secreted proteins may act as cytokines or may affect
CC   cellular proliferation or differentiation or may act as immune system
CC   regulators, haematopoiesis regulators, tissue growth regulators,
CC   regulators of reproductive hormones or cell movement or have
CC   chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or tumour
CC   inhibition activity. The DNAs can be used in forensic procedures to
CC   identify individuals or in diagnostic procedures to identify individuals
CC   having genetic diseases resulting from abnormal expression of the genes
CC   corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 47 BP; 10 A; 24 C; 2 G; 11 U; 0 other;

  Query Match           100.0%;  Score 24;  DB 20;  Length 47;
  Best Local Similarity  100.0%;  Pred. No. 0.13;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          ||||||||||||||||||||||||
Db      2 gcauccuacucccauccaauucca 25


Search completed: June  2, 2002, 20:08:48
Job time: 14733 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 15:53:00 ; Search time 2378.87 Seconds
(without alignments)
211.124 Million cell updates/sec

Title: US-09-663-600A-2_COPY_1_24
Perfect score: 24
Sequence: 1 gcauccuacucccauccaauucca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24	100.0	46	6	AR177314	AR177314 Sequence
2	24	100.0	47	6	AR177313	AR177313 Sequence
3	24	100.0	65	6	A58296	A58296 Sequence 1
4	23	95.8	46	6	A58299	A58299 Sequence 4
5	23	95.8	46	6	A58300	A58300 Sequence 5
6	19.2	80.0	196356	2	AC096552	AC096552 Homo sapi
c 7	19.2	80.0	336028	2	AC092096	AC092096 Mus muscu
c 8	18.8	78.3	47295	8	F22M8	AC020622 Sequence
c 9	18.8	78.3	78249	8	AP004529	AP004529 Lotus jap
c 10	18.8	78.3	100849	9	HUMNEUROF	L05367 Human oligo
11	18.8	78.3	142898	2	AC095140	AC095140 Rattus no
c 12	18.8	78.3	200768	2	AC084272	AC084272 Mus muscu
13	18.8	78.3	203504	2	AC068162	AC068162 Homo sapi
14	18.8	78.3	204991	2	AC068980	AC068980 Homo sapi
15	18.8	78.3	204993	2	AC024447	AC024447 Homo sapi
16	18.8	78.3	205825	2	AC092942	AC092942 Homo sapi
c 17	18.8	78.3	297898	9	AC004526	AC004526 Homo sapi
c 18	18.4	76.7	146722	2	AL139254	AL139254 Homo sapi
19	18.4	76.7	245341	2	AC073436	AC073436 Mus muscu
c 20	18.2	75.8	1312	10	AF016040	AF016040 Mus muscu
21	18.2	75.8	1489	4	AB004272	AB004272 Bos tauru
c 22	18.2	75.8	6635	8	D14487	D14487 Lentinus ed
c 23	18.2	75.8	8172	10	AF017112	AF017112 Mus muscu
24	18.2	75.8	73605	9	AL136309	AL136309 Human DNA
c 25	18.2	75.8	73778	6	AX344575	AX344575 Sequence
c 26	18.2	75.8	112054	2	AC068730	AC068730 Homo sapi
c 27	18.2	75.8	115154	2	AP003804	AP003804 Oryza sat
28	18.2	75.8	136299	9	HS359N14	Z97206 Human DNA s
c 29	18.2	75.8	153615	2	AL645665	AL645665 Homo sapi
c 30	18.2	75.8	155523	2	AC025367	AC025367 Homo sapi
c 31	18.2	75.8	160054	2	AP003179	AP003179 Homo sapi
c 32	18.2	75.8	167214	2	AC099216	AC099216 Rattus no
33	18.2	75.8	167263	2	AC063966	AC063966 Homo sapi
34	18.2	75.8	169635	2	HSAC001235	AC001235 Homo sapi
35	18.2	75.8	190694	2	AC098056	AC098056 Rattus no
c 36	18.2	75.8	193029	2	AL355594	AL355594 Homo sapi
c 37	18.2	75.8	208123	2	AL662894	AL662894 Mus muscu
c 38	18.2	75.8	349980	6	AX344574	AX344574 Sequence
39	17.8	74.2	1867	3	DROPERIODJ	L07827 Drosophila
40	17.8	74.2	1867	3	DROPERIODM	L07824 Drosophila
41	17.8	74.2	1867	3	DROPERIODO	L07822 Drosophila
42	17.8	74.2	1867	3	DROPERIODQ	L07820 Drosophila
43	17.8	74.2	1870	3	DROPERIOC	L07831 Drosophila
44	17.8	74.2	1870	3	DROPERIODI	L07828 Drosophila
45	17.8	74.2	1870	3	DROPERIODZ	L07811 Drosophila

ALIGNMENTS

RESULT 1
AR177314
LOCUS AR177314 46 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6312922.
ACCESSION AR177314
VERSION AR177314.1 GI:17919669
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 46)
AUTHORS Edwards,J.-B.Dumas.Milne., Duclert,A. and Bougueleret,L.
TITLE Complementary DNAs
JOURNAL Patent: US 6312922-A 2 06-NOV-2001;
FEATURES Location/Qualifiers
source 1. .46
/organism="unknown"
BASE COUNT 10 a 24 c 1 g 11 t
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 46;
Best Local Similarity 75.0%; Pred. No. 0.33;

```
  Matches   18;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          |||:||:||:||||:||||::|||
Db      1 GCATCCTACTCCCATCCAATTCCA 24


RESULT   2
AR177313
LOCUS       AR177313                  47 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 1 from patent US 6312922.
ACCESSION   AR177313
VERSION     AR177313.1  GI:17919668
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 47)
  AUTHORS   Edwards,J.-B.Dumas.Milne., Duclert,A. and Bougueleret,L.
  TITLE     Complementary DNAs
  JOURNAL   Patent: US 6312922-A 1 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .47
                     /organism="unknown"
BASE COUNT       10 a      24 c       2 g      11 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 6;  Length 47;
  Best Local Similarity   75.0%;  Pred. No. 0.33;
  Matches   18;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          |||:||:||:||||:||||::|||
Db      2 GCATCCTACTCCCATCCAATTCCA 25


RESULT   3
A58296
LOCUS       A58296                    65 bp    DNA     linear   PAT 05-MAR-1998
DEFINITION  Sequence 1 from Patent WO9634981.
ACCESSION   A58296
VERSION     A58296.1  GI:3713960
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 65)
  AUTHORS   Nicolaevna,M.I. and Dumas,M.E.
  TITLE     METHOD FOR THE SPECIFIC COUPLING OF THE CAP OF THE EXTREMITY 5' OF
            A FRAGMENT mRNA AND PREPARATION OF mRNA AND COMPLETE cDNA
  JOURNAL   Patent: WO 9634981-A 1 07-NOV-1996;
            GENSET (FR)
COMMENT     Other publication AU 5982996 961121
            Other publication FR 2733765 961108
            Other publication FR 2733762 961108.
FEATURES             Location/Qualifiers
     source          1. .65
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       17 a      29 c       2 g      17 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 6;  Length 65;
  Best Local Similarity   75.0%;  Pred. No. 0.33;
  Matches   18;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          |||:||:||:||||:||||::|||
Db     20 GCATCCTACTCCCATCCAATTCCA 43


RESULT   4
A58299
LOCUS       A58299                    46 bp    DNA     linear   PAT 05-MAR-1998
DEFINITION  Sequence 4 from Patent WO9634981.
ACCESSION   A58299
VERSION     A58299.1  GI:3713963
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Nicolaevna,M.I. and Dumas,M.E.
  TITLE     METHOD FOR THE SPECIFIC COUPLING OF THE CAP OF THE EXTREMITY 5' OF
            A FRAGMENT mRNA AND PREPARATION OF mRNA AND COMPLETE cDNA
  JOURNAL   Patent: WO 9634981-A 4 07-NOV-1996;
            GENSET (FR)
COMMENT     Other publication AU 5982996 961121
            Other publication FR 2733765 961108
            Other publication FR 2733762 961108.
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       10 a      24 c       0 g      11 t       1 others
ORIGIN


  Query Match             95.8%;  Score 23;  DB 6;  Length 46;
  Best Local Similarity   73.9%;  Pred. No. 1;
  Matches   17;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          ||:||:||:||||:||||::|||
Db      2 CATCCTACTCCCATCCAATTCCA 24


RESULT   5
A58300
LOCUS       A58300                    46 bp    DNA     linear   PAT 05-MAR-1998
DEFINITION  Sequence 5 from Patent WO9634981.
ACCESSION   A58300
VERSION     A58300.1  GI:3713964
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Nicolaevna,M.I. and Dumas,M.E.
  TITLE     METHOD FOR THE SPECIFIC COUPLING OF THE CAP OF THE EXTREMITY 5' OF
            A FRAGMENT mRNA AND PREPARATION OF mRNA AND COMPLETE cDNA
  JOURNAL   Patent: WO 9634981-A 5 07-NOV-1996;
            GENSET (FR)
COMMENT     Other publication AU 5982996 961121
            Other publication FR 2733765 961108
            Other publication FR 2733762 961108.
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       10 a      24 c       0 g      11 t       1 others
ORIGIN


  Query Match             95.8%;  Score 23;  DB 6;  Length 46;
  Best Local Similarity   73.9%;  Pred. No. 1;
  Matches   17;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucca 24
          ||:||:||:||||:||||::|||
Db      2 CATCCTACTCCCATCCAATTCCA 24


RESULT   6
```

```
AC096552
LOCUS       AC096552              196356 bp    DNA     linear   HTG 13-DEC-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-41M22, WORKING DRAFT SEQUENCE,
            8 unordered pieces.
ACCESSION   AC096552 AC027552
VERSION     AC096552.2  GI:17647039
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196356)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196356)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Dec 13, 2001 this sequence version replaced gi:15638726.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0041M22
            Drafting center: WIBR
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 29%
            Sequencing vector: plasmid; 65%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 94% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 193421 bases at least Q40
            Consensus quality: 194158 bases at least Q30
            Consensus quality: 194505 bases at least Q20
            Insert size: 196000; agarose-fp
            Insert size: 195193; sum-of-contigs
            Quality coverage: 12.65 in Q20 bases; agarose-fp
            Quality coverage: 12.67 in Q20 bases; sum-of-contigs
            ------------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1966: contig of 1966 bp in length
            *     1967     2066: gap of unknown length
            *     2067     3358: contig of 1292 bp in length
            *     3359     3458: gap of unknown length
            *     3459    12020: contig of 8562 bp in length
            *    12021    12120: gap of unknown length
            *    12121    32759: contig of 20639 bp in length
            *    32760    32859: gap of unknown length
            *    32860   192962: contig of 160103 bp in length
            *   192963   193062: gap of unknown length
            *   193063   193525: contig of 463 bp in length
            *   193526   193625: gap of unknown length
            *   193626   194887: contig of 1262 bp in length
            *   194888   194987: gap of unknown length
            *   194988   196356: contig of 1369 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196356
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-41M22"
     misc_feature    1. .1966
                     /note="assembly_name:Contig10"
     misc_feature    2067. .3358
                     /note="assembly_name:Contig11"
     misc_feature    3459. .12020
                     /note="assembly_name:Contig12
                     clone_end:T7
                     vector_side:left"
     misc_feature    12121. .32759
                     /note="assembly_name:Contig13"
     misc_feature    32860. .192962
                     /note="assembly_name:Contig14"
     misc_feature    193063. .193525
                     /note="assembly_name:Contig3"
     misc_feature    193626. .194887
                     /note="assembly_name:Contig8"
     misc_feature    194988. .196356
                     /note="assembly_name:Contig9"
BASE COUNT    57531 a  35141 c  38098 g  64879 t    707 others
ORIGIN


  Query Match             80.0%;  Score 19.2;  DB 2;  Length 196356;
  Best Local Similarity   62.5%;  Pred. No. 64;
  Matches   15;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcauccuacucccauccaauucca 24
          |||: |:||: |||:||||::| |
Db 188328 GCATACTACTACCATCCAATTCTA 188351


RESULT   7
AC092096/c
LOCUS       AC092096              336028 bp    DNA     linear   HTG 05-FEB-2002
DEFINITION  Mus musculus chromosome 10 clone rp23-39k4, WORKING DRAFT SEQUENCE,
            19 unordered pieces.
ACCESSION   AC092096
VERSION     AC092096.17  GI:18497142
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 336028)
  AUTHORS   Do,T. and Roe,B.A.
  TITLE     Mus musculus Chromosome 10 BAC Clone rp23-39k4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 336028)
  AUTHORS   Do,T. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2001) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:18201826.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3074: contig of 3074 bp in length
            *     3075     3174: gap of unknown length
            *     3175     6014: contig of 2840 bp in length
            *     6015     6114: gap of unknown length
```

```
            *     6115     9345: contig of 3231 bp in length
            *     9346     9445: gap of unknown length
            *     9446    12107: contig of 2662 bp in length
            *    12108    12207: gap of unknown length
            *    12208    18756: contig of 6549 bp in length
            *    18757    18856: gap of unknown length
            *    18857    23733: contig of 4877 bp in length
            *    23734    23833: gap of unknown length
            *    23834    30446: contig of 6613 bp in length
            *    30447    30546: gap of unknown length
            *    30547    37867: contig of 7321 bp in length
            *    37868    37967: gap of unknown length
            *    37968    45708: contig of 7741 bp in length
            *    45709    45808: gap of unknown length
            *    45809    53387: contig of 7579 bp in length
            *    53388    53487: gap of unknown length
            *    53488    63026: contig of 9539 bp in length
            *    63027    63126: gap of unknown length
            *    63127    74700: contig of 11574 bp in length
            *    74701    74800: gap of unknown length
            *    74801    85536: contig of 10736 bp in length
            *    85537    85636: gap of unknown length
            *    85637   108073: contig of 22437 bp in length
            *   108074   108173: gap of unknown length
            *   108174   137882: contig of 29709 bp in length
            *   137883   137982: gap of unknown length
            *   137983   183433: contig of 45451 bp in length
            *   183434   183533: gap of unknown length
            *   183534   222828: contig of 39295 bp in length
            *   222829   222928: gap of unknown length
            *   222929   262269: contig of 39341 bp in length
            *   262270   262369: gap of unknown length
            *   262370   336028: contig of 73659 bp in length.
FEATURES             Location/Qualifiers
     source          1. .336028
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /clone="rp23-39k4"
                     /clone_lib="rp23"
BASE COUNT   101266 a  67078 c  66482 g  99351 t   1851 others
ORIGIN


  Query Match             80.0%;  Score 19.2;  DB 2;  Length 336028;
  Best Local Similarity   66.7%;  Pred. No. 63;
  Matches   16;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcauccuacucccauccaauucca 24
           |||: |:||:||| :|||| :|||
Db 326247 GCATTCTACTCCCTTCCAAATCCA 326224


RESULT   8
F22M8/c
LOCUS       F22M8                   47295 bp    DNA     linear   PLN 05-JAN-2001
DEFINITION  Sequence of BAC F22M8 from Arabidopsis thaliana chromosome 1,
            complete sequence.
ACCESSION   AC020622
VERSION     AC020622.3  GI:12039253
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 47295)
  AUTHORS   Liu,S.X., Sakano,H., Yu,G., Lee,J.M., Lenz,C., Pham,P., Toriumi,M.,
            Chin,C., Chiou,J., Choi,E., Chung,M., Gonzalez,A., Howng,B.,
            Liu,A., Vaysberg,M., Alfalfi,H., Brooks,S., Buehler,E., Chao,Q.,
            Conn,L., Conway,A.B., Hansen,N.F., Johnson-Hopson,C., Khan,S.,
            Kim,C., Lam,B., Miranda,M., Nguen,M., Palm,C.J., Shinn,P.,
            Southwick,A., Davis,R.W., Ecker,J.R., Federspiel,N.A. and
            Theologis,A.
  TITLE     The sequence of BAC F22M8 from Arabidopsis thaliana chromosome 1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 47295)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-2000) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   3  (bases 1 to 47295)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   4  (bases 1 to 47295)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2000) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   5  (bases 1 to 47295)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     On Jan 5, 2001 this sequence version replaced gi:6693373.
            The sequence is of BAC F22M8 from Arabidopsis thaliana chromosome
            1.  In order to facilitate the joining of overlapping clones in the
            future for creation of larger contigs, we provide overlaps between
            overlapping sumbitted clones.  The 3' end of this sequence overlaps
            by 2482 bp the 5' end of the sequence of the clone T7I23.
FEATURES             Location/Qualifiers
     source          1. .47295
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F22M8"
     gene            complement(<375. .1476)
                     /gene="F22M8.1"
     CDS             complement(join(<375. .459,548. .643,748. .903,986. .1135,
                     1246. .1476))
                     /gene="F22M8.1"
                     /note="Identical to dimethyladenosine transferase (PFC1)
                     from Arabidopsis thaliana gb|AF051326 and contains a
                     Ribosomal RNA Adenine Dimethylases PF|00398 domain.  This
                     gene may be cut off."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF76480.1"
                     /db_xref="GI:8570453"
                     /translation="MMNAVITSATINCNSLSPSWTCGDNSPSKLLLGEISAALSRRRT
                     VKVSCGKSSPDDYHSTLKSLNSRGRFPRKSLGQHYMLNSDINDQLASAADVKEGDFVL
                     EIGPGTGSLTNVLINLGATVLAIEKDPHMVDLVSERFAGSDKFKVLQEDFVKCHIRSH
                     MLSILETRRLSHPDSALAKVVSNLPFNISTDVVKLLLPMGDIFSKVVLLLQDEAALRL
                     VEPALRTSEYRPINILINFYS"
     tRNA            1585. .1657
                     /note="codon recognized: GUA"
                     /product="tRNA-Val"
     gene            complement(1759. .4192)
                     /gene="F22M8.2"
     CDS             complement(join(1759. .2001,2094. .2195,2502. .2560,
                     2642. .3006,3090. .3271,3349. .3519,3598. .3774,3998. .4192))
                     /gene="F22M8.2"
                     /note="Contains similarity to excision repair protein
                     ERCC5 from Homo sapiens gi|1082359 and contains XPG
                     N-terminal PF|00752 and XPG I-region PF|00867 domains."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF76467.1"
                     /db_xref="GI:8570440"
                     /translation="MGVGGNFWDLLRPYAQQQGFDFLRNKRVAVDLSFWIVQHETAVK
                     GFVLKPHLRLTFFRTINLFSKFGAYPVFVVDGTPSPLKSQARISRFFRSSGIDTCNLP
                     VIKDGVSVERNKLFSEWVRECVELLELLGIPVLKANGEAEALCAQLNSQGFVDACITP
                     DSDAFLFGAMCVIKDIKPNSREPFECYHMSHIESGLGLKRKHLIAISLLVGNDYDSGG
```

```
                VLGIGVDKALRIVREFSEDQVLERLQDIGNGLQPAVPGGIKSGDDGEEFRSEMKKRSP
                HCSRCGHLGSKRTHFKSSCEHCGCDSGCIKKPLGFRCECSFCSKDRDLREQKKTNDWW
                IKVCDKIALAPEFPNRKIIELYLSDGLMTGDGSSMSWGTPDTGMLVDLMVQNDNGDCF
                LLTDECIGLVQSAFPDETEHFLHEKKLRESKKKNVSEEETATPRATTMGVQRSITDFY
                RSAKKAAAGQSIETGGSSKASAEKKRQATSTSSSNLTKSVRRRLLFG"
tRNA            complement(4476. .4548)
                /note="codon recognized: AAG"
                /product="tRNA-Lys"
gene            5587. .8212
                /gene="F22M8.3"
CDS             join(5587. .6182,6538. .8212)
                /gene="F22M8.3"
                /note="Contains similarity to p69d gene from Lycopersicon
                esculentum gb|Y17278 and contains a Peptidase S8 PF|00082
                domain."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAF76468.1"
                /db_xref="GI:8570441"
                /translation="MFFRSFIVFFFLIFFASNVSSRKQTYVIHTVTTSTKHIVTSLFN
                SLQTENINDDDFSLPEIHYIYENAMSGFSATLTDDQLDTVKNTKGFISAYPDELLSLH
                TTYSHEFLGLEFGIGLWNETSLSSDVIIGLVDTGISPEHVSFRDTHMTPVPSRWRGSC
                DEGTNFSSSECNKKIIGASAFYKGYESIVGKINETTDFRSTRDAQGHGTHTASTAAGD
                IVPKANYFGQAKGLASGMRFTSRIAAYKACWALGCASTDVIAAIDRAILDGVDVISLS
                LGGSSRPFYVDPIAIAGFGAMQKNIFVSCSAGNSGPTASTVSNGAPWLMTVAASYTDR
                TFPAIVRIGNRKSLVGSSLYKGKSLKNLPLAFNRTAGEESGAVFCIRDSLKRELVEGK
                IVICLRGASGRTAKGEEVKRSGGAAMLLVSTEAEGEELLADPHVLPAVSLGFSDGKTL
                LNYLAGAANATASVRFRGTAYGATAPMVAAFSSRGPSVAGPEIAKPDIAAPGLNILAG
                WSPFSSPSLLRSDPRRVQFNIISGTSMACPHISGIAALIKSVHGDWSPAMIKSAIMTT
                ARITDNRNRPIGDRGAAGAESAATAFAFGAGNVDPTRAVDPGLVYDTSTVDYLNYLCS
                LNYTSERILLFSGTNYTCASNAVVLSPGDLNYPSFAVNLVNGANLKTVRYKRTVTNVG
                SPTCEYMVHVEEPKGVKVRVEPKVLKFQKARERLSYTVTYDAEASRNSSSSSFGVLVW
                ICDKYNVRSPIAVTWE"
gene            complement(8796. .11032)
                /gene="F22M8.4"
CDS             complement(join(8796. .8960,9081. .9196,9288. .9426,
                9666. .9752,9853. .10003,10092. .10216,10778. .11032))
                /gene="F22M8.4"
                /note="Contains similarity to arsenite translocating
                ATPase from Homo sapiens gb|AF047469 and contains a 4Fe-4S
                iron sulfur cluster binding protein PF|00142 domain.  EST
                gb|N37510 comes from this gene."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAF76469.1"
                /db_xref="GI:8570442"
                /translation="MAADLPEATVQNILDQESLKWVFVGGKGGVGKTTCSSILAICLA
                SVRSSVLIISTDPAHNLSDAFQQRFTKSPTLVQGFSNLFAMEVDPTVETDDMAGTDGM
                DGLFSDLANAIPGIDEAMSFAEMLKLVQTMDYATIVFDTAPTGHTLRLLQFPATLEKG
                LSKLMSLKSRFGGLMTQMSRMFGMEDEFGEDALLGRLEGLKDVIEQDMTTFVCVCIPE
                FLSLYETERLVQELAKFEIDTHNIIINQVLYDDEDVESKLLRARMRMQQKYLDQFYML
                YDDFNITKLPLLPEEVTGVEALKAFSHKFLTPYHPTTSRSNVEELERKVHTLRLQLKT
                AEEELERVKSG"
gene            11405. .14708
                /gene="F22M8.5"
CDS             join(11405. .11458,11545. .11641,11740. .11914,11993. .12091,
                12368. .12394,12495. .12595,12754. .12843,12966. .13046,
                13180. .13241,13346. .13402;13497. .13634,14187. .14302,
                14405. .14497,14573. .14708)
                /gene="F22M8.5"
                /note="Contains similarity to a hypothetical protein
                gi|6983878 from Oryza Sativa BAC gb|AP001168."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAF76470.1"
                /db_xref="GI:8570443"
                /translation="MAISEEEAKLERFLDWLQVNGGELRGCNIKYSDSLKGFGIFAST
                STQASDEVLLVVPLDLAITPMRVLQDPLLGPECQKMFEQGQVDDRFLMILFLTLERLR
                INSSWKPYLDMLPTRFGNPLWFSDDDILELKGTNLYHATDKVSFEHFLWANSVFWSRA
                LNIPLPHSFVFPQSQDDTGECTSTKIQAQPAPSVGSGDTIWVEGLVPGIDFCNHDLKP
                VATWEVDGIGSVSRVPFSMYLLSVAQRPIPKKEISISYGNKGNEELLYLYGFVIDNNP
                DDYLMIKEMLVNFVLTSVVTFNNGFIQVHYPVEAIPSIPFSDSKGQLLEAQLVESRQG
                EQPSETEVRMAVWEACGDSGALQLLVDLLNSKMMKLEENSGTEEQDARLLEEACVLES
                HEESRDLDGRRMSRNKWSSVVYRRGQKQLTRLLLKEAEHALHLALSSDH"
gene            complement(15242. .18010)
                /gene="F22M8.6"
CDS             complement(join(15242. .15346,15453. .15497,15581. .15658,
                15970. .16623,16952. .17066,17159. .17607,17714. .18010))
                /gene="F22M8.6"
                /note="Contains similarity to an unnamed protein from Homo
                sapiens gb|AK001277 and contains an Ank repeat PF|00023
                domain."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAF76471.1"
                /db_xref="GI:8570444"
                /translation="MATTGAAATTEIGNPEYKQPRSIFDLTADFFDSCRLSNPSETQS
                AFGPPKTFDPEEAEEDKSSKDGVILDRWTCNTCKIEFLSLQDQRYHFKSDIHRLNIKL
                SVAGKAILKEEDVDELTSESVQDYDVSSISGSEDEAETRPPSFHFDAQKGIDKKKLFF
                RLQSGDKVSIWKCLIMDDAESVSFENDRGVSVDCCGSLVENEVTERLRNLIRENKDDR
                QMRVVLLASGGHFAGTVFNGKSVVAHKTFHRYVVRAKAGKKQSTKDASGRSIHSAGAS
                LRRYNELALKKDIQELLASWKPYFDGAACVFVHAPSSSRQLLFNGGKPYFSSQNCAVR
                NVPFTIRRPTFKESQRIYNQLTQIAHVTEEIFVNRPEVTKANTVVQTHNEDSGKTSRK
                EEPDETSSSNIILEEPNRIEEDIEDGVTGTSTALHEAAKSGDCERVMEFLEEGMDPCA
                KDERGRTPYMLANEKEVRNTFRRFMALNLEKWNWHDAKVPSPLTKEMEESQAAKQAEK
                DAKQKARTKELKKLRKAREKKAQAEAAQAEKEKPISKVEEVRRAMAAQREKRAAAAER
                RMASLNIQSSSSTS"
gene            18283. .19920
                /gene="F22M8.7"
CDS             join(18283. .18285,18409. .18483,18880. .19041,19459. .19617,
                19837. .19920)
                /gene="F22M8.7"
                /note="Contains similarity to cyclophilin E from
                Dictyostelium discoideum gb|AF215865 and contains a
                Cyclophilin type peptidyl-prolyl cis-trans isomerase
                PF|00160 domain."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAF76472.1"

  Query Match             78.3%;  Score 18.8;  DB 8;  Length 47295;
  Best Local Similarity   68.2%;  Pred. No. 1e+02;
  Matches   15;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucc 23
          ||:||: |:||||:|||| :||
Db  14594 CATCCTTCTCCCATCCAAGTCC 14573


RESULT   9
AP004529/c
LOCUS       AP004529               78249 bp    DNA     linear   PLN 14-DEC-2001
DEFINITION  Lotus japonicus genomic DNA, chromosome 1, clone:LjT01K12, TM0058a,
            complete sequence.
ACCESSION   AP004529
VERSION     AP004529.1  GI:17736896
KEYWORDS    HTG.
SOURCE      Lotus japonicus DNA, clone_lib:LjT library clone:LjT01K12.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (sites)
  AUTHORS   Sato,S., Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. I. Sequence
            Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb
            Regions of the Genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78249)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:81-438-52-3935,
```

```
           Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .78249
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT01K12"
                     /clone_lib="LjT library"
                     /note="TM0058a, a part of TAC clone:TM0058"
BASE COUNT     25249 a  14277 c  13703 g  25020 t
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 8;  Length 78249;
  Best Local Similarity   63.6%;  Pred. No. 1e+02;
  Matches   14;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 cauccuacucccauccaauucc 23
           ||:||:||:||| :| ||::||
Db  18457 CATCCTACTCCCTTCAAATTCC 18436


RESULT  10
HUMNEUROF/c
LOCUS       HUMNEUROF              100849 bp    DNA     linear   PRI 20-SEP-1995
DEFINITION  Human oligodendrocyte myelin glycoprotein (OMG) exons 1-2;
            neurofibromatosis 1 (NF1) exons 28-49; ecotropic viral integration
            site 2B (EVI2B) exons 1-2; ecotropic viral integration site 2A
            (EVI2A) exons 1-2; adenylate kinase (AK3) exons 1-2.
ACCESSION   L05367 L03723
VERSION     L05367.1  GI:189152
KEYWORDS    adenylate kinase; ecotropic viral integration site 2A; ecotropic
            viral integration site 2B; neurofibromatosis type 1;
            oligodendrocyte myelin glycoprotein.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100849)
  AUTHORS   Cawthon,R.M., Weiss,R.B., Xu,G., Viskochil,D., Culver,M.,
            Stevens,J., Robertson,M., Dunn,D., Gesteland,R., O'Connell,P. and
            White,R.
  TITLE     A major segment of the neurofibromatosis type 1 gene: cDNA
            sequence, genomic structure, and point mutations [published erratum
            appears in Cell 1990 Aug 10;62(3):following 608]
  JOURNAL   Cell 62 (1), 193-201 (1990)
  MEDLINE   90304909
REFERENCE   2  (bases 1 to 100849)
  AUTHORS   Cawthon,R.M., O'Connell,P., Buchberg,A.M., Viskochil,D.,
            Weiss,R.B., Culver,M., Stevens,J., Jenkins,N.A., Copeland,N.G. and
            White,R.
  TITLE     Identification and characterization of transcripts from the
            neurofibromatosis 1 region: the sequence and genomic structure of
            EVI2 and mapping of other transcripts
  JOURNAL   Genomics 7 (4), 555-565 (1990)
  MEDLINE   90353953
REFERENCE   3  (bases 1 to 100849)
  AUTHORS   Cawthon,R.M., Andersen,L.B., Buchberg,A.M., Xu,G.F., O'Connell,P.,
            Viskochil,D., Weiss,R.B., Wallace,M.R., Marchuk,D.A.,
            Culver,M.Stevens.J., Jenkins,N.A., Copeland,N.G., Collins,F.S. and
            White,R.
  TITLE     cDNA sequence and genomic structure of EV12B, a gene lying within
            an intron of the neurofibromatosis type 1 gene
  JOURNAL   Genomics 9 (3), 446-460 (1991)
  MEDLINE   91236164
REFERENCE   4  (bases 1 to 100849)
  AUTHORS   Viskochil,D., Cawthon,R., O'Connell,P., Xu,G.F., Stevens,J.,
            Culver,M., Carey,J. and White,R.
  TITLE     The gene encoding the oligodendrocyte-myelin glycoprotein is
            embedded within the neurofibromatosis type 1 gene
  JOURNAL   Mol. Cell. Biol. 11 (2), 906-912 (1991)
  MEDLINE   91117257
REFERENCE   5  (bases 1 to 100849)
  AUTHORS   Wallace,M.R., Andersen,L., Letcher,R., Odeh,H., Saulino,A.M.,
            Fountain,J., Brereton,A., Nicholson,J., Mitchell,A.,
            Brownstein,B.H. and Collins,F.
  TITLE     A de novo Alu insertion results in neurofibromatosis type 1
  JOURNAL   Nature 353 (6347), 864-866 (1991)
  MEDLINE   92049738
REFERENCE   6  (bases 1 to 100849)
  AUTHORS   Weiss,R.B., Dunn,D., DiSera,L., Wheatley,W., Kimball,A., Rote,C.,
            Cherry,J., Duval,B., Lee,R., Ferguson,M.W.J. and Gesteland,R.F.
  TITLE     The Human Neurofibromatosis Type 1 locus: genomic sequence of the
            3' region
  JOURNAL   Unpublished (1992)
REFERENCE   7  (bases 1 to 100849)
  AUTHORS   Xu,G., O'Connell,P., Stevens,J. and White,R.
  TITLE     Characterization of human adenylate kinase 3 (AK3) cDNA and mapping
            of the AK3 pseudogene to an intron of the NF1 gene
  JOURNAL   Genomics 13 (3), 537-542 (1992)
  MEDLINE   92347846
COMMENT     Submitting author Robert Weiss may be contacted at the following
            address and telephone number:                          e-mail:
            bob@corona.med.utah.edu                                  postal: Room
            7160, Eccles Institute of Human Genetics, University of Utah, Salt
            Lake City, Utah 84112          telephone: (801) 585-3436.
FEATURES             Location/Qualifiers
     source          1. .100849
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="17q11.2"
     intron          <1. .36919
                     /number=1
     gene            complement(5761. .8342)
                     /gene="OMG"
     exon            complement(5761. .7447)
                     /gene="OMG"
                     /note="G00-127-563"
                     /citation=[4]
                     /number=2
     intron          complement(7448. .8265)
                     /gene="OMG"
                     /note="G00-127-563"
                     /number=1
     exon            complement(8266. .8342)
                     /gene="OMG"
                     /note="G00-127-563"
                     /citation=[4]
                     /number=1
     repeat_region   12224. .12233
                     /rpt_family="GT"
     exon            complement(14658. .16737)
                     /gene="EVI2B"
                     /note="G00-125-901"
                     /citation=[3]
                     /number=2
     gene            complement(14658. .25133)
                     /gene="EVI2B"
     intron          complement(16738. .25076)
                     /gene="EVI2B"
                     /note="G00-125-901"
                     /number=1
     repeat_region   24830. .24856
                     /rpt_family="GT"
     exon            complement(25077. .25133)
                     /gene="EVI2B"
                     /note="G00-125-901"
                     /citation=[3]
                     /number=1
     repeat_region   28465. .28503
                     /rpt_family="AC"
     exon            complement(28542. .30121)
                     /gene="EVI2A"
                     /note="G00-125-191"
                     /citation=[2]
                     /number=2
```

```
gene            complement(28542. .32788)
                /gene="EVI2A"
intron          complement(30122. .32588)
                /gene="EVI2A"
                /note="G00-125-191"
                /number=1
exon            complement(32589. .32788)
                /gene="EVI2A"
                /note="G00-125-191"
                /citation=[2]
                /number=1
repeat_region   33193. .33232
                /rpt_family="AC"
exon            36920. .37340
                /gene="NF1"
                /note="36920 = nt. 5032 from full-length neurofibromin
                cDNA (HUMNF1B,M89914); G00-120-231"
                /number=28
gene            36920. .86010
                /gene="NF1"
intron          37341. .38596
                /gene="NF1"
                /note="G00-120-231"
                /number=28
exon            38597. .38937
                /gene="NF1"
                /note="G00-120-231"
                /number=29
intron          38938. .41643
                /gene="NF1"
                /note="G00-120-231"
                /number=29
exon            41644. .41846
                /gene="NF1"
                /note="G00-120-231"
                /number=30
intron          41847. .46170
                /gene="NF1"
                /note="G00-120-231"
                /number=30
exon            46171. .46364
                /gene="NF1"
                /note="G00-120-231"
                /number=31
variation       46216
                /gene="NF1"
                /note="leu to pro change, exon 31; G00-120-231"
                /citation=[1]
                /replace="c"
variation       46260
                /gene="NF1"
                /note="arg to stop change, exon 31; G00-120-231"
                /citation=[1]
                /replace="t"
intron          46365. .47772
                /gene="NF1"
                /note="G00-120-231"
                /number=31
variation       46600. .46740
                /gene="NF1"
                /note="t17:22 translocation breakpoint; G00-120-231"
                /replace=""
exon            47773. .47913
                /gene="NF1"
                /note="G00-120-231"
                /number=32
intron          47914. .48075
                /gene="NF1"
                /note="G00-120-231"
                /number=32
variation       48030
                /gene="NF1"
                /note="de novo ALU insertion; G00-120-231"
                /citation=[5]
                /replace=""
exon            48076. .48355
                /gene="NF1"
                /note="G00-120-231"
```

```
  Query Match             78.3%;  Score 18.8;  DB 9;  Length 100849;
  Best Local Similarity   72.7%;  Pred. No. 1e+02;
  Matches   16;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 auccuacucccauccaauucca 24
           |:||:|| |||| ||||::||| 
Db   99895 ATCCTACCCCCACCCAATTCCA 99874


RESULT  11
AC095140
LOCUS       AC095140              142898 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-7O21, *** SEQUENCING IN PROGRESS ***,
            59 unordered pieces.
ACCESSION   AC095140
VERSION     AC095140.2  GI:17942101
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 142898)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142898)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Dec 20, 2001 this sequence version replaced gi:15625694.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GCSI
              Center clone name: CH230-7O21
          -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329First call to
          findPhrapList
              Consensus quality: 112126 bases at least Q40
              Consensus quality: 120850 bases at least Q30
              Consensus quality: 128435 bases at least Q20
              Estimated insert size: 118743; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 1.5x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 59 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     4143: contig of 4143 bp in length
          *     4144     4243: gap of unknown length
          *     4244     9274: contig of 5031 bp in length
          *     9275     9374: gap of unknown length
          *     9375    15325: contig of 5951 bp in length
          *    15326    15425: gap of unknown length
          *    15426    22207: contig of 6782 bp in length
          *    22208    22307: gap of unknown length
          *    22308    25695: contig of 3388 bp in length
          *    25696    25795: gap of unknown length
          *    25796    31599: contig of 5804 bp in length
          *    31600    31699: gap of unknown length
          *    31700    35413: contig of 3714 bp in length
          *    35414    35513: gap of unknown length
          *    35514    39478: contig of 3965 bp in length
          *    39479    39578: gap of unknown length
          *    39579    43076: contig of 3498 bp in length
          *    43077    43176: gap of unknown length
          *    43177    45443: contig of 2267 bp in length
          *    45444    45543: gap of unknown length
          *    45544    48215: contig of 2672 bp in length
          *    48216    48315: gap of unknown length
          *    48316    51300: contig of 2985 bp in length
          *    51301    51400: gap of unknown length
          *    51401    53322: contig of 1922 bp in length
          *    53323    53422: gap of unknown length
          *    53423    56005: contig of 2583 bp in length
          *    56006    56105: gap of unknown length
          *    56106    58776: contig of 2671 bp in length
          *    58777    58876: gap of unknown length
          *    58877    60821: contig of 1945 bp in length
          *    60822    60921: gap of unknown length
          *    60922    62970: contig of 2049 bp in length
          *    62971    63070: gap of unknown length
          *    63071    64702: contig of 1632 bp in length
          *    64703    64802: gap of unknown length
          *    64803    67527: contig of 2725 bp in length
          *    67528    67627: gap of unknown length
          *    67628    71110: contig of 3483 bp in length
          *    71111    71210: gap of unknown length
          *    71211    74105: contig of 2895 bp in length
          *    74106    74205: gap of unknown length
          *    74206    75907: contig of 1702 bp in length
          *    75908    76007: gap of unknown length
          *    76008    79435: contig of 3428 bp in length
          *    79436    79535: gap of unknown length
          *    79536    82171: contig of 2636 bp in length
          *    82172    82271: gap of unknown length
          *    82272    84362: contig of 2091 bp in length
          *    84363    84462: gap of unknown length
          *    84463    86374: contig of 1912 bp in length
          *    86375    86474: gap of unknown length
          *    86475    88823: contig of 2349 bp in length
          *    88824    88923: gap of unknown length
          *    88924    90852: contig of 1929 bp in length
          *    90853    90952: gap of unknown length
          *    90953    93358: contig of 2406 bp in length
          *    93359    93458: gap of unknown length
          *    93459    94986: contig of 1528 bp in length
          *    94987    95086: gap of unknown length
          *    95087    96893: contig of 1807 bp in length
          *    96894    96993: gap of unknown length
          *    96994    98840: contig of 1847 bp in length
          *    98841    98940: gap of unknown length
          *    98941   100977: contig of 2037 bp in length
          *   100978   101077: gap of unknown length
          *   101078   102920: contig of 1843 bp in length
          *   102921   103020: gap of unknown length
          *   103021   104528: contig of 1508 bp in length
          *   104529   104628: gap of unknown length
          *   104629   106895: contig of 2267 bp in length
          *   106896   106995: gap of unknown length
          *   106996   108325: contig of 1330 bp in length
          *   108326   108425: gap of unknown length
          *   108426   109446: contig of 1021 bp in length
          *   109447   109546: gap of unknown length
          *   109547   111148: contig of 1602 bp in length
          *   111149   111248: gap of unknown length
          *   111249   113082: contig of 1834 bp in length
          *   113083   113182: gap of unknown length
          *   113183   114745: contig of 1563 bp in length
          *   114746   114845: gap of unknown length
          *   114846   116015: contig of 1170 bp in length
          *   116016   116115: gap of unknown length
          *   116116   118467: contig of 2352 bp in length
          *   118468   118567: gap of unknown length
          *   118568   120547: contig of 1980 bp in length
          *   120548   120647: gap of unknown length
          *   120648   122231: contig of 1584 bp in length
          *   122232   122331: gap of unknown length
          *   122332   123520: contig of 1189 bp in length
          *   123521   123620: gap of unknown length
          *   123621   125501: contig of 1881 bp in length
          *   125502   125601: gap of unknown length
          *   125602   126912: contig of 1311 bp in length
          *   126913   127012: gap of unknown length
          *   127013   128420: contig of 1408 bp in length
          *   128421   128520: gap of unknown length
          *   128521   129699: contig of 1179 bp in length
          *   129700   129799: gap of unknown length
          *   129800   130901: contig of 1102 bp in length
          *   130902   131001: gap of unknown length
          *   131002   132646: contig of 1645 bp in length
          *   132647   132746: gap of unknown length
          *   132747   134251: contig of 1505 bp in length
          *   134252   134351: gap of unknown length
          *   134352   135499: contig of 1148 bp in length
          *   135500   135599: gap of unknown length
          *   135600   137135: contig of 1536 bp in length
          *   137136   137235: gap of unknown length
          *   137236   138470: contig of 1235 bp in length
          *   138471   138570: gap of unknown length

Query Match              78.3%;  Score 18.8;  DB 2;  Length 142898;
```

```
  Best Local Similarity   63.6%;  Pred. No. 1e+02;
  Matches   14;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucc 23
          ||:||: |:||||:||| ::||
Db 107853 CATCCTTCTCCCATCCAGTTCC 107874


RESULT  12
AC084272/c
LOCUS       AC084272              200768 bp    DNA     linear   HTG 29-JAN-2002
DEFINITION  Mus musculus chromosome 16 clone rp23-11g21, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC084272
VERSION     AC084272.20  GI:18390233
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200768)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE     Mus musculus Chromosome 16 BAC Clone rp23-11g21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200768)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jan 29, 2002 this sequence version replaced gi:18139381.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3097: contig of 3097 bp in length
            *     3098     3197: gap of unknown length
            *     3198     7898: contig of 4701 bp in length
            *     7899     7998: gap of unknown length
            *     7999    17599: contig of 9601 bp in length
            *    17600    17699: gap of unknown length
            *    17700    36839: contig of 19140 bp in length
            *    36840    36939: gap of unknown length
            *    36940    68422: contig of 31483 bp in length
            *    68423    68522: gap of unknown length
            *    68523   111403: contig of 42881 bp in length
            *   111404   111503: gap of unknown length
            *   111504   149778: contig of 38275 bp in length
            *   149779   149878: gap of unknown length
            *   149879   200768: contig of 50890 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200768
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="rp23-11g21"
                     /clone_lib="rp23"
BASE COUNT    56087 a  46789 c  45810 g  51371 t    711 others
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 2;  Length 200768;
  Best Local Similarity   63.6%;  Pred. No. 1e+02;
  Matches   14;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucc 23
          ||:||:| :||| :||||::||
Db 189911 CATCCTAATCCCTTCCAATTCC 189890


RESULT  13
AC068162
LOCUS       AC068162              203504 bp    DNA     linear   HTG 09-JUN-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-543P4 map 3, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC068162
VERSION     AC068162.2  GI:8389502
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 203504)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-543P4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203504)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 9, 2000 this sequence version replaced gi:7670156.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10146
                Center clone name: 543_P_4
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 184552 bases at least Q40
                Consensus quality: 194359 bases at least Q30
                Consensus quality: 198180 bases at least Q20
                Insert size: 208000; agarose-fp
                Insert size: 200504; sum-of-contigs
                Quality coverage: 3.8 in Q20 bases; agarose-fp
                Quality coverage: 4.0 in Q20 bases; sum-of-contigs
            ----------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
```

```
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1303: contig of 1303 bp in length
*     1304 1403: gap of      100 bp
*     1404     2468: contig of 1065 bp in length
*     2469 2568: gap of      100 bp
*     2569     3769: contig of 1201 bp in length
*     3770 3869: gap of      100 bp
*     3870     5347: contig of 1478 bp in length
*     5348 5447: gap of      100 bp
*     5448     6492: contig of 1045 bp in length
*     6493 6592: gap of      100 bp
*     6593     8593: contig of 2001 bp in length
*     8594 8693: gap of      100 bp
*     8694    10995: contig of 2302 bp in length
*    10996 11095: gap of      100 bp
*    11096    13032: contig of 1937 bp in length
*    13033 13132: gap of      100 bp
*    13133    17251: contig of 4119 bp in length
*    17252 17351: gap of      100 bp
*    17352    20481: contig of 3130 bp in length
*    20482 20581: gap of      100 bp
*    20582    23646: contig of 3065 bp in length
*    23647 23746: gap of      100 bp
*    23747    27501: contig of 3755 bp in length
*    27502 27601: gap of      100 bp
*    27602    33325: contig of 5724 bp in length
*    33326 33425: gap of      100 bp
*    33426    37218: contig of 3793 bp in length
*    37219 37318: gap of      100 bp
*    37319    42731: contig of 5413 bp in length
*    42732 42831: gap of      100 bp
*    42832    46787: contig of 3956 bp in length
*    46788 46887: gap of      100 bp
*    46888    53272: contig of 6385 bp in length
*    53273 53372: gap of      100 bp
*    53373    61153: contig of 7781 bp in length
*    61154 61253: gap of      100 bp
*    61254    68270: contig of 7017 bp in length
*    68271 68370: gap of      100 bp
*    68371    77795: contig of 9425 bp in length
*    77796 77895: gap of      100 bp
*    77896    85280: contig of 7385 bp in length
*    85281 85380: gap of      100 bp
*    85381    94950: contig of 9570 bp in length
*    94951 95050: gap of      100 bp
*    95051   103176: contig of 8126 bp in length
*   103177 103276: gap of      100 bp
*   103277   110866: contig of 7590 bp in length
*   110867 110966: gap of      100 bp
*   110967   120234: contig of 9268 bp in length
*   120235 120334: gap of      100 bp
*   120335   129289: contig of 8955 bp in length
*   129290 129389: gap of      100 bp
*   129390   139564: contig of 10175 bp in length
*   139565 139664: gap of      100 bp
*   139665   150765: contig of 11101 bp in length
*   150766 150865: gap of      100 bp
*   150866   167813: contig of 16948 bp in length
*   167814 167913: gap of      100 bp
*   167914   185753: contig of 17840 bp in length
*   185754 185853: gap of      100 bp
*   185854   203504: contig of 17651 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203504
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-543P4"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1303
                     /note="assembly_fragment"
     misc_feature    1404. .2468
                     /note="assembly_fragment"
     misc_feature    2569. .3769
                     /note="assembly_fragment"
     misc_feature    3870. .5347
                     /note="assembly_fragment"
     misc_feature    5448. .6492
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    6593. .8593
                     /note="assembly_fragment"
     misc_feature    8694. .10995
                     /note="assembly_fragment"
     misc_feature    11096. .13032
                     /note="assembly_fragment"
     misc_feature    13133. .17251
                     /note="assembly_fragment"
     misc_feature    17352. .20481
                     /note="assembly_fragment"
     misc_feature    20582. .23646
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    23747. .27501
                     /note="assembly_fragment"
     misc_feature    27602. .33325
                     /note="assembly_fragment"
     misc_feature    33426. .37218
                     /note="assembly_fragment"
     misc_feature    37319. .42731
                     /note="assembly_fragment"
     misc_feature    42832. .46787
                     /note="assembly_fragment"
     misc_feature    46888. .53272
                     /note="assembly_fragment"
     misc_feature    53373. .61153
                     /note="assembly_fragment"
     misc_feature    61254. .68270
                     /note="assembly_fragment"
     misc_feature    68371. .77795
                     /note="assembly_fragment"
     misc_feature    77896. .85280
                     /note="assembly_fragment"
     misc_feature    85381. .94950
                     /note="assembly_fragment"
     misc_feature    95051. .103176
                     /note="assembly_fragment"
     misc_feature    103277. .110866
                     /note="assembly_fragment"
     misc_feature    110967. .120234
                     /note="assembly_fragment"
     misc_feature    120335. .129289
                     /note="assembly_fragment"
     misc_feature    129390. .139564
                     /note="assembly_fragment"
     misc_feature    139665. .150765
                     /note="assembly_fragment"

  Query Match           78.3%;  Score 18.8;  DB 2;  Length 203504;
  Best Local Similarity   68.2%;  Pred. No. 1e+02;
  Matches   15;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 cauccuacucccauccaauucc 23
           || ||:||:||||:||| ::||
Db   65837 CAACCTACTCCCATCCACTTCC 65858


RESULT  14
```

```
AC068980
LOCUS       AC068980              204991 bp    DNA     linear   HTG 05-OCT-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-552A14, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC068980
VERSION     AC068980.10  GI:15741365
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 204991)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204991)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 25, 2001 this sequence version replaced gi:11128156.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBGY
                Center clone name: RP11-552A14
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 14% of reads
                Chemistry: Dye-terminator Big Dye: 86% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 196487 bases at least Q40
                Consensus quality: 203036 bases at least Q30
                Consensus quality: 206264 bases at least Q20
                Estimated insert size: 205706; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    52522: contig of 52522 bp in length
            *    52523    52622: gap of unknown length
            *    52623    91359: contig of 38737 bp in length
            *    91360    91459: gap of unknown length
            *    91460   125166: contig of 33707 bp in length
            *   125167   125266: gap of unknown length
            *   125267   152897: contig of 27631 bp in length
            *   152898   152997: gap of unknown length
            *   152998   173601: contig of 20604 bp in length
            *   173602   173701: gap of unknown length
            *   173702   188647: contig of 14946 bp in length
            *   188648   188747: gap of unknown length
            *   188748   202570: contig of 13823 bp in length
            *   202571   202670: gap of unknown length
            *   202671   204991: contig of 2321 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-552A14"
BASE COUNT    60973 a  40617 c  40715 g  61970 t    716 others
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 2;  Length 204991;
  Best Local Similarity   68.2%;  Pred. No. 1e+02;
  Matches   15;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucc 23
          || ||:||:||||:||| ::||
Db  45115 CAACCTACTCCCATCCACTTCC 45136


RESULT  15
AC024447
LOCUS       AC024447              204993 bp    DNA     linear   HTG 11-APR-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-479J2 map 3, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC024447
VERSION     AC024447.2  GI:7534015
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 204993)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-479J2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204993)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
```

```
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 11, 2000 this sequence version replaced gi:7108243.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6997
                Center clone name: 479_J_2
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 184405 bases at least Q40
                Consensus quality: 195348 bases at least Q30
                Consensus quality: 199562 bases at least Q20
                Insert size: 202493; sum-of-contigs
                Quality coverage: 3.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1059: contig of 1059 bp in length
            *     1060 1159: gap of      100 bp
            *     1160     3098: contig of 1939 bp in length
            *     3099 3198: gap of      100 bp
            *     3199     4151: contig of 953 bp in length
            *     4152 4251: gap of      100 bp
            *     4252     6077: contig of 1826 bp in length
            *     6078 6177: gap of      100 bp
            *     6178    10158: contig of 3981 bp in length
            *    10159 10258: gap of      100 bp
            *    10259    13440: contig of 3182 bp in length
            *    13441 13540: gap of      100 bp
            *    13541    16429: contig of 2889 bp in length
            *    16430 16529: gap of      100 bp
            *    16530    19606: contig of 3077 bp in length
            *    19607 19706: gap of      100 bp
            *    19707    24259: contig of 4553 bp in length
            *    24260 24359: gap of      100 bp
            *    24360    28314: contig of 3955 bp in length
            *    28315 28414: gap of      100 bp
            *    28415    32584: contig of 4170 bp in length
            *    32585 32684: gap of      100 bp
            *    32685    37793: contig of 5109 bp in length
            *    37794 37893: gap of      100 bp
            *    37894    44505: contig of 6612 bp in length
            *    44506 44605: gap of      100 bp
            *    44606    50170: contig of 5565 bp in length
            *    50171 50270: gap of      100 bp
            *    50271    57212: contig of 6942 bp in length
            *    57213 57312: gap of      100 bp
            *    57313    68003: contig of 10691 bp in length
            *    68004 68103: gap of      100 bp
            *    68104    77484: contig of 9381 bp in length
            *    77485 77584: gap of      100 bp
            *    77585    86863: contig of 9279 bp in length
            *    86864 86963: gap of      100 bp
            *    86964    96495: contig of 9532 bp in length
            *    96496 96595: gap of      100 bp
            *    96596   107929: contig of 11334 bp in length
            *   107930 108029: gap of      100 bp
            *   108030   116856: contig of 8827 bp in length
            *   116857 116956: gap of      100 bp
            *   116957   128225: contig of 11269 bp in length
            *   128226 128325: gap of      100 bp
            *   128326   142566: contig of 14241 bp in length
            *   142567 142666: gap of      100 bp
            *   142667   160140: contig of 17474 bp in length
            *   160141 160240: gap of      100 bp
            *   160241   179301: contig of 19061 bp in length
            *   179302 179401: gap of      100 bp
            *   179402   204993: contig of 25592 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204993
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-479J2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1059
                     /note="assembly_fragment"
     misc_feature    1160. .3098
                     /note="assembly_fragment"
     misc_feature    3199. .4151
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    4252. .6077
                     /note="assembly_fragment"
     misc_feature    6178. .10158
                     /note="assembly_fragment"
     misc_feature    10259. .13440
                     /note="assembly_fragment"
     misc_feature    13541. .16429
                     /note="assembly_fragment"
     misc_feature    16530. .19606
                     /note="assembly_fragment"
     misc_feature    19707. .24259
                     /note="assembly_fragment"
     misc_feature    24360. .28314
                     /note="assembly_fragment"
     misc_feature    28415. .32584
                     /note="assembly_fragment"
     misc_feature    32685. .37793
                     /note="assembly_fragment"
     misc_feature    37894. .44505
                     /note="assembly_fragment"
     misc_feature    44606. .50170
                     /note="assembly_fragment"
     misc_feature    50271. .57212
                     /note="assembly_fragment"
     misc_feature    57313. .68003
                     /note="assembly_fragment"
     misc_feature    68104. .77484
                     /note="assembly_fragment"
     misc_feature    77585. .86863
                     /note="assembly_fragment"
```

```
     misc_feature    86964. .96495
                     /note="assembly_fragment"
     misc_feature    96596. .107929
                     /note="assembly_fragment"
     misc_feature    108030. .116856
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    116957. .128225
                     /note="assembly_fragment"
     misc_feature    128326. .142566
                     /note="assembly_fragment"
     misc_feature    142667. .160140
                     /note="assembly_fragment"
     misc_feature    160241. .179301
                     /note="assembly_fragment"
     misc_feature    179402. .204993
                     /note="assembly_fragment"
BASE COUNT    62968 a  39652 c  40375 g  59489 t   2509 others
ORIGIN


  Query Match            78.3%;  Score 18.8;  DB 2;  Length 204993;
  Best Local Similarity  68.2%;  Pred. No. 1e+02;
  Matches   15;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cauccuacucccauccaauucc 23
          || ||:||:||||:||| ::||
Db  81605 CAACCTACTCCCATCCACTTCC 81626


Search completed: June  2, 2002, 18:57:16
Job time: 11056 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 19:59:57 ; Search time 3664.46 Seconds
(without alignments)
2541.411 Million cell updates/sec

Title: US-09-663-600A-139_COPY_36_725
Perfect score: 690
Sequence: 1 atggcctctcttggcctcca.....acagcctgacagggtatgtg 690

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	610	88.4	822	10	BG386074	BG386074 602455248
2	599.6	86.9	728	10	BG325755	BG325755 602424466
3	585	84.8	680	10	BE304667	BE304667 601105783
4	579.4	84.0	978	10	BG164062	BG164062 602341087
5	568.8	82.4	945	10	BE513091	BE513091 601171545
6	547.6	79.4	2829	11	AK004990	AK004990 Mus muscu
7	503.6	73.0	727	10	BI101652	BI101652 602887470
8	502	72.8	724	10	BI100253	BI100253 602885716
9	492	71.3	643	10	BG328625	BG328625 602427889
10	484	70.1	777	10	BI102100	BI102100 602887847
11	473.4	68.6	758	10	BI102679	BI102679 602888236
12	460.8	66.8	706	10	BI148156	BI148156 602912333
13	460	66.7	494	10	BG385562	BG385562 602453858
14	456.8	66.2	846	10	BF789255	BF789255 602105156
15	451.8	65.5	750	9	AW475316	AW475316 un64d02.y
16	437.6	63.4	1114	10	BF789836	BF789836 602105395
17	434.6	63.0	673	9	AI746723	AI746723 ul06f09.y
18	426.2	61.8	754	10	BG972599	BG972599 602841188
19	424	61.4	787	10	BF781962	BF781962 602106220
20	422.8	61.3	804	10	BF784189	BF784189 602108052
21	422.4	61.2	704	10	BI330754	BI330754 602982419
22	414.4	60.1	659	9	AI789490	AI789490 uk99h03.y
23	406.8	59.0	674	10	BI102491	BI102491 602889685
24	402.2	58.3	856	10	BF541534	BF541534 602067906
25	394.8	57.2	569	10	BE667438	BE667438 151799 MA
26	389.6	56.5	886	10	BF032123	BF032123 601559727
27	381	55.2	628	9	AI788398	AI788398 uk58c09.y
28	378.8	54.9	892	10	BF781303	BF781303 602106979
29	361.4	52.4	555	9	AV604013	AV604013 AV604013
c 30	358.2	51.9	560	9	AI953481	AI953481 wq29b02.x
31	354.4	51.4	593	9	AI789119	AI789119 uk51c01.y
32	353	51.2	1030	10	BF788818	BF788818 602110526
33	348.8	50.6	808	10	BG469035	BG469035 602510779
34	347	50.3	770	9	AI173400	AI173400 ue91f05.y
35	344	49.9	551	9	AV605969	AV605969 AV605969
36	328.6	47.6	616	9	AW475430	AW475430 un66c07.y
37	323.6	46.9	403	9	AI116661	AI116661 ue28a02.y
38	322.6	46.8	514	9	AI788419	AI788419 uk58e09.y
39	313.6	45.4	517	10	BE480970	BE480970 166244 BA
40	307.8	44.6	611	10	BF785226	BF785226 602108423
41	294.8	42.7	757	10	BI143586	BI143586 602907418
42	291.8	42.3	520	9	AI746647	AI746647 ul01g04.y
43	291.6	42.3	816	10	BI144992	BI144992 602909649
44	288.6	41.8	376	9	AW610976	AW610976 un74h05.y
45	283.2	41.0	361	9	AI048112	AI048112 ud67a10.y

ALIGNMENTS

RESULT 1
BG386074
LOCUS BG386074 822 bp mRNA linear EST 12-MAR-2001
DEFINITION 602455248F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583675 5', mRNA sequence.
ACCESSION BG386074
VERSION BG386074.1 GI:13279520
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1307 row: p column: 12
High quality sequence stop: 746.
FEATURES Location/Qualifiers
source 1. .822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583675"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in

```
                    the laboratory of Gerald M. Rubin (University of
                    California, Berkeley) using  ZAP-cDNA synthesis kit
                    (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      181 a    224 c    217 g    200 t
ORIGIN


  Query Match             88.4%;  Score 610;  DB 10;  Length 822;
  Best Local Similarity   95.7%;  Pred. No. 3.8e-158;
  Matches  660;  Conservative    0;  Mismatches   25;  Indels    5;  Gaps    3;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 186

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 246

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 306

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     307 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 366

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     367 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 426

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     427 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 486

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     487 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 546

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     547 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 606

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     607 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 666

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           |||||||||||||||||||||||||||||||| |||||||||||||||||| |  |
Db     667 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCG-TCCAACTACTACGATGCCTACAAGCCA 725

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
             |  |   |  ||||||||||||||| ||||  |  ||||  || ||||||||||||||
Db     726 AACTC---TTGCACAAGGAGCTCTCCCAGGCTGGTCCAACTCCCAAAAGTCAAGAGTGAG 782

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
           || | | |||||  ||||| || | ||||
Db     783 GTCCAAT-CTACAAGCTGACGGGTTGTGTG 811


RESULT   2
BG325755
LOCUS       BG325755                728 bp    mRNA    linear   EST 27-FEB-2001
DEFINITION  602424466F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4562273 5',
            mRNA sequence.
ACCESSION   BG325755
VERSION     BG325755.1  GI:13132179
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1274  row: d  column: 18
            High quality sequence stop: 708.
FEATURES             Location/Qualifiers
     source          1. .728
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4562273"
                     /clone_lib="NIH_MGC_14"
                     /tissue_type="renal cell adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      154 a    205 c    183 g    186 t
ORIGIN


  Query Match             86.9%;  Score 599.6;  DB 10;  Length 728;
  Best Local Similarity   97.5%;  Pred. No. 2.7e-155;
  Matches  673;  Conservative    0;  Mismatches    9;  Indels    8;  Gaps    6;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      30 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGC-TTTGGGC 88

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 148

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     149 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 208

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 268

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 328

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 388

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 448

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 508

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     509 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 568
```

```
Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    569 CTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTACGATG-CTACCAAGCC 627

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||    | ||||||||||||||||||||    |   | ||||||||||||||||||||
Db    628 CAAC--TCTTGCACAAGGAGCTCTCCAAGGC--TGGTCAACTCCCAAAGTCAAGAGTGAG 683

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          ||||||||||||| ||||||||||||||||
Db    684 -TCAATTCCTACAG-CTGACAGGGTATGTG 711


RESULT   3
BE304667
LOCUS       BE304667                680 bp    mRNA    linear   EST 13-JUL-2000
DEFINITION  601105783F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:2988309 5',
            mRNA sequence.
ACCESSION   BE304667
VERSION     BE304667.1  GI:9176036
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM77  row: f  column: 22
            High quality sequence start: 27
            High quality sequence stop: 680.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2988309"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      146 a    189 c    182 g    163 t
ORIGIN


  Query Match             84.8%;  Score 585;  DB 10;  Length 680;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-151;
  Matches  585;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    106 gtcggtgccagcattgtgacagcagttggcttctccaagggcctctggatggaatgtgcc 165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38 GTCGGTGCCAGCATTGTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCC 97

Qy    166 acacacagcacaggcatcacccagtgtgacatctatagcacccttctgggcctgcccgct 225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     98 ACACACAGCACAGGCATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCT 157

Qy    226 gacatccaggctgcccaggccatgatggtgacatccagtgcaatctcctccctggcctgc 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 GACATCCAGGCTGCCCAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGC 217

Qy    286 attatctctgtggtgggcatgagatgcacagtcttctgccaggaatcccgagccaaagac 345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 ATTATCTCTGTGGTGGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGAC 277

Qy    346 agagtggcggtagcaggtggagtctttttcatccttggaggcctcctgggattcattcct 405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 AGAGTGGCGGTAGCAGGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCT 337

Qy    406 gttgcctggaatcttcatgggatcctacgggacttctactcaccactggtgcctgacagc 465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    338 GTTGCCTGGAATCTTCATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGC 397

Qy    466 atgaaatttgagattggagaggctctttacttgggcattatttcttccctgttctccctg 525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    398 ATGAAATTTGAGATTGGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTG 457

Qy    526 atagctggaatcatcctctgcttttcctgctcatcccagagaaatcgctccaactactac 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    458 ATAGCTGGAATCATCCTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACTACTAC 517

Qy    586 gatgcctaccaagcccaacctcttgccacaaggagctctccaaggcctggtcaacctccc 645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    518 GATGCCTACCAAGCCCAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCC 577

Qy    646 aaagtcaagagtgagttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||||||||||||||||||
Db    578 AAAGTCAAGAGTGAGTTCAATTCCTACAGCCTGACAGGGTATGTG 622


RESULT   4
BG164062
LOCUS       BG164062                978 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602341087F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4449212 5',
            mRNA sequence.
ACCESSION   BG164062
VERSION     BG164062.1  GI:12670765
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 978)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10232  row: m  column: 21
            High quality sequence stop: 689.
FEATURES             Location/Qualifiers
     source          1. .978
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4449212"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
```

```
BASE COUNT      220 a    279 c    276 g    203 t
ORIGIN

  Query Match             84.0%;  Score 579.4;  DB 10;  Length 978;
  Best Local Similarity   94.8%;  Pred. No. 1.2e-149;
  Matches  655;  Conservative    0;  Mismatches   26;  Indels   10;  Gaps    5;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggctttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      25 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGC-TTTGGGC 83

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 143

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     144 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 203

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     204 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 263

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     264 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 323

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     324 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 383

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     384 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 443

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     444 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 503

Qy     481 ggagaggctc-tttacttgggcattatttcttccctgttctccctgatagctggaatcat 539
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     504 GGAGAGGCTCTTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCAT 563

Qy     540 cctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagc 599
           |||||||||||||||||  ||    ||      |||||||||||||||||| |
Db     564 CCTCTGCTTTTCCTGCTCAATCCCAGAGAAATCGCTCAACTACTACGATGCCTACAA--- 620

Qy     600 ccaacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtga 659
               |  ||   || ||||||||||||||  |||||||||||| || |||||||| ||
Db     621 --GCCAACTCTGCAAAAGGAGCTCTCCAAG--CTGGTCAACCTCCAAAGGTCAAGAG-GA 675

Qy     660 gttcaattcctacagcctgacagggtatgtg 690
           ||||||||||||||| | ||||||||||||||
Db     676 GTTCAATTCCTACAGGCCGACAGGGTATGTG 706


RESULT   5
BE513091
LOCUS       BE513091                945 bp    mRNA    linear   EST 07-AUG-2000
DEFINITION  601171545F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:3545171 5',
            mRNA sequence.
ACCESSION   BE513091
VERSION     BE513091.1  GI:9720302
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)

COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM241  row: i  column: 12
            High quality sequence start: 36
            High quality sequence stop: 782.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3545171"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      194 a    262 c    263 g    226 t
ORIGIN

  Query Match             82.4%;  Score 568.8;  DB 10;  Length 945;
  Best Local Similarity   93.0%;  Pred. No. 1e-146;
  Matches  640;  Conservative    0;  Mismatches   42;  Indels    6;  Gaps    4;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggctttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 246

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 306

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     307 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 366

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     367 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 426

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     427 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 486

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     487 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 546

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     547 GGTGGAGTC-TTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 605

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttga--ga 478
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db     606 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAAGAT 665

Qy     479 ttggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatca 538
           | | ||||||||||||||||||  ||  | ||   |  ||||||||||||||||||||
Db     666 TGGCAGAGGCTCTTTACTTGGGCTTATTTCTTCCCTGTTCTCCCCTGATAGCTGGAATCA 725

Qy     539 -tcctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaa 597
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
```

```
Db    726 TTCCTCTGCTTTTCCTGCTCATCCCAGAGAAATCGGTCCAACTACTACGATGCCTACCAA 785

Qy    598 gcccaacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagt 657
          | |    |   | |   |||||||||| ||| || |   |||| || ||| |||||||
Db    786 GGCGCCACCTCTTGCCACAGGAGCTCTCCCAGGGCTTGGTCACCTTCCCAAGGCAAGAGT 845

Qy    658 gagttcaattcctacagcctgacagggt 685
            | |  ||||||||||||||||||||||
Db    846 --GGTTCATTCCTACAGCCTGACAGGGT 871


RESULT   6
AK004990
LOCUS       AK004990               2829 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male liver cDNA, RIKEN full-length enriched
            library, clone:1300013G06:claudin 2, full insert sequence.
ACCESSION   AK004990
VERSION     AK004990.1  GI:12836594
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1300013G06.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2829)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .2829
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1904439"
                     /db_xref="taxon:10090"
                     /clone="1300013G06"
                     /sex="male"
                     /tissue_type="liver"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     gene            153. .845
                     /gene="Cldn2"
     CDS             153. .845
                     /gene="Cldn2"
                     /note="claudin 2
                     data source:MGD, source key:MGI:1276110, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB23725.1"
                     /db_xref="GI:12836595"
                     /translation="MASLGVQLVGYILGLLGLLGTSIAMLLPNWRTSSYVGASIVTAV
                     GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAMSSLACIISVVGM
                     RCTVFCQDSRAKDRVAVVGGVFFILGGILGFIPVAWNLHGILRDFYSPLVPDSMKFEI
                     GEALYLGIISALFSLVAGVILCFSCSPQGNRTNYYDGYQAQPLATRSSPRSAQQPKAK
                     SEFNSYSLTGYV"
     polyA_signal    2808. .2813
                     /note="putative"
     polyA_site      2829
                     /note="putative"
BASE COUNT      718 a    707 c    674 g    730 t
ORIGIN


  Query Match             79.4%;  Score 547.6;  DB 11;  Length 2829;
  Best Local Similarity   87.1%;  Pred. No. 1.3e-140;
  Matches  601;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||| |||||| ||||||| ||||||||||||||||||||| ||||||| || |||
Db    153 ATGGCCTCCCTTGGCGTCCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 212

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
            |||   |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db    213 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 272

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| ||||||||||||||||||| ||||| ||||||||||||||||
Db    273 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 332

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
```

```
Db     333 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 392

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           |||||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db     393 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 452

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db     453 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 512

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           |||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db     513 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 572

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||| ||||| |||||||||||||| || ||||| ||||||||||||||||||||||||
Db     573 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGAGATT 632

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           |||||||||| |||||||||| || ||  ||||||| ||  || |||| ||| |||||
Db     633 GGAGAGGCTCTGTACTTGGGCATCATCTCAGCCCTGTTTTCTTTGGTAGCCGGAGTCATC 692

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           || |||||||||||||  ||||| | |||||  |||||||||| |||| |||||| |||
Db     693 CTTTGCTTTTCCTGCTCGCCCCAGGGCAATCGTACCAACTACTATGATGGCTACCAGGCC 752

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
           || |||||||||| ||||||||||||  ||| |||||  ||||||| ||||||||||
Db     753 CAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGCTCAACAGCCCAAAGCCAAGAGTGAG 812

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
           ||||| || |||||||||| |||||||||
Db     813 TTCAACTCATACAGCCTGACTGGGTATGTG 842


RESULT   7
BI101652
LOCUS       BI101652                727 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602887470F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5042825
            5', mRNA sequence.
ACCESSION   BI101652
VERSION     BI101652.1  GI:14552545
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11117  row: k  column: 18
            High quality sequence stop: 722.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5042825"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      142 a    196 c    195 g    194 t
ORIGIN


  Query Match             73.0%;  Score 503.6;  DB 10;  Length 727;
  Best Local Similarity   86.8%;  Pred. No. 1.1e-128;
  Matches  554;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           |||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db      90 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 149

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db     150 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 209

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||| || |||||||| |||||||||||||||||| ||||| ||||||||||||||||
Db     210 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 269

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           |||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db     270 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 329

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           |||||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db     330 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 389

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db     390 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 449

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           |||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db     450 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 509

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||| ||||| |||||||||||||| || ||||| ||||||||||||||||||||||||
Db     510 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGAGATT 569

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           |||||||||| |||||||||| || ||  ||||||| ||  || |||| ||| |||||
Db     570 GGAGAGGCTCTGTACTTGGGCATCATCTCAGCCCTGTTTTCTTTGGTAGCCGGAGTCATC 629

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           || |||||||||||||  ||||| | |||||  |||||||||| |||| |||||| |||
Db     630 CTTTGCTTTTCCTGCTCGCCCCAGGGCAATCGTACCAACTACTATGATGGCTACCAGGCC 689

Qy     601 caacctcttgccacaaggagctctccaaggcctggtca 638
           || |||||||||| ||||||||||||||  ||| |||
Db     690 CAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGCTCA 727


RESULT   8
BI100253
LOCUS       BI100253                724 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602885716F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5041370
            5', mRNA sequence.
ACCESSION   BI100253
VERSION     BI100253.1  GI:14551146
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
```

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11113  row: o  column: 03
           High quality sequence stop: 722.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5041370"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      136 a    194 c    198 g    195 t      1 others
ORIGIN


  Query Match            72.8%;  Score 502;  DB 10;  Length 724;
  Best Local Similarity  86.4%;  Pred. No. 2.9e-128;
  Matches  589;  Conservative    0;  Mismatches   90;  Indels    3;  Gaps    3;

Qy     11 ttggcctccaacttgtgggctacatcctaggccttctggggcttttgggcacactggttg 70
          | ||| | ||||| |||||||||||||||||||||| ||||||| || ||||||    |||
Db     37 TGGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGCACATCCATTG 96

Qy     71 ccatgctgctccccagct-ggaaaacaagttcttatgtcggtgccagcattgtgacagca 129
          |||||||||| |||| || ||  ||| ||||| ||||| |||||||||| |||||| ||
Db     97 CCATGCTGCTTCCCAACTNGGCGAACGAGTTCCTATGTTGGTGCCAGCA-TGTGACGGCG 155

Qy    130 gttggcttctccaagggcctctggatggaatgtgccacacacagcacaggcatcacccag 189
          |||||||| ||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db    156 GTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGCATCACCCAG 215

Qy    190 tgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcccaggccatg 249
          || || ||||| || |||||| | || || || ||||||||||||||||||||||||||
Db    216 TGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCCCAGGCCATG 275

Qy    250 atggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtgggcatgaga 309
          ||||||| |||||||||| ||||| ||||| || ||||||||||||||||||||||||
Db    276 ATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTGGGCATGAGA 335

Qy    310 tgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagcaggtggagtc 369
          ||||| || |||||||||||| || ||||| || |||||||||| ||||  |||||||||
Db    336 TGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTGGGTGGAGTC 395

Qy    370 tttttcatccttggaggcctcctgggattcattcctgttgcctggaatcttcatgggatc 429
          |||||||||||||| ||| ||||||| || || || ||||| |||||||||||||| |||
Db    396 TTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTTCATGGCATC 455

Qy    430 ctacgggacttctactcaccactggtgcctgacagcatgaaatttgagattggagaggct 489
          || |||||||||||||| || ||||| |||||||||||||||| ||||||||||||||||
Db    456 CTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTCGAGATTGGAGAGGCT 515

Qy    490 ctttacttgggcattatttcttccctgttctccctgatagctggaatcatcctctgcttt 549
          || |||||||||| || ||  ||||||| ||  || |||| ||| ||||||| ||||||
Db    516 CTGTACTTGGGCATCATCTCAGCCCTGTTTTCTTTGGTAGCCGGAGTCATCCTTTGCTTT 575

Qy    550 tcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagccca-acctct 608
          ||||||||  ||||| | |||||  |||||||||| |||| |||||| ||||| ||||||
Db    576 TCCTGCTCGCCCCAGGGCAATCGTACCAACTACTATGATGGCTACCAGGCCCAGACCTCT 635

Qy    609 tgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgagttcaattc 668
          |||||| ||||||||||||||  ||| |||||  ||||||| ||||||||||||||| ||
Db    636 TGCCACTAGGAGCTCTCCAAGATCTGCTCAACAGCCCAAAGCCAAGAGTGAGTTCAACTC 695

Qy    669 ctacagcctgacagggtatgtg 690
          ||||||||||| |||||||||
Db    696 ATACAGCCTGACTGGGTATGTG 717


RESULT   9
BG328625
LOCUS       BG328625                643 bp    mRNA    linear   EST 27-FEB-2001
DEFINITION  602427889F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4547230 5',
            mRNA sequence.
ACCESSION   BG328625
VERSION     BG328625.1  GI:13135063
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1235  row: a  column: 23
            High quality sequence stop: 641.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4547230"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      130 a    173 c    177 g    163 t
ORIGIN


  Query Match            71.3%;  Score 492;  DB 10;  Length 643;
  Best Local Similarity  99.4%;  Pred. No. 1.6e-125;
  Matches  525;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    116 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 175

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagc-at 119
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    176 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCGAT 235

Qy    120 tgtgacagcagttggcttctccaagggcctctggatggaatgtgccaca-cacagcacag 178
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    236 TGTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACAGCACAGCACAG 295

Qy    179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    296 GCATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTG 355

Qy    239 cccaggccatgatggtgacatccagtgcaatctcctccctggcctgca-ttatctctgtg 297
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
```

```
Db    356 CCCAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCAGTTATCTCTGTG 415

Qy    298 gtgggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggta 357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 GTGGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTA 475

Qy    358 gcaggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaat 417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 GCAGGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAAT 535

Qy    418 cttcatgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgag 477
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    536 CTTCATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAG 595

Qy    478 attggagaggctctttacttgggcattatttcttccctgttctccctg 525
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    596 ATTGGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTG 643


RESULT  10
BI102100
LOCUS       BI102100                777 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602887847F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043251
            5', mRNA sequence.
ACCESSION   BI102100
VERSION     BI102100.1  GI:14552993
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11118  row: m  column: 12
            High quality sequence stop: 749.
FEATURES             Location/Qualifiers
     source          1. .777
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043251"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      166 a    201 c    210 g    200 t
ORIGIN


  Query Match             70.1%;  Score 484;  DB 10;  Length 777;
  Best Local Similarity   86.3%;  Pred. No. 2.9e-123;
  Matches  535;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||| |||||| | ||||| ||||||||||||||||||||| ||||||| || |||
Db    153 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 212

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db    213 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 272

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| ||||||||||||||||||||| ||||| ||||||||||||||||
Db    273 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 332

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db    333 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 392

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          |||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db    393 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 452

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db    453 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 512

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          |||||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db    513 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 572

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||| ||||| |||||||||||||| || ||||| |||||||||||||||||||||||||
Db    573 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGAGATT 632

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          |||||||||||| |||||||||||| || ||  ||||||   |  || |||| ||| |||||
Db    633 GGAGAGGCTCTGTACTTGGGCATCATCTCAGCCCTGTATACTATGGTAGCCGGAGTCATC 692

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          || |||||||||||||||  ||||| | |||||  |||||||||| |||| ||| || |||
Db    693 CTTTGCTTTTCCTGCTCGCCCCAGGGCAATCGTACCAACTACTATGATGGCTAACAGGCC 752

Qy    601 caacctcttgccacaaggag 620
          || | |||||||||| |||||
Db    753 CAGCTTCTTGCCACTAGGAG 772


RESULT  11
BI102679
LOCUS       BI102679                758 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602888236F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5043489
            5', mRNA sequence.
ACCESSION   BI102679
VERSION     BI102679.1  GI:14553572
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 758)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11119  row: g  column: 10
            High quality sequence stop: 756.
FEATURES             Location/Qualifiers
     source          1. .758
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5043489"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
```

```
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      149 a    205 c    207 g    197 t
ORIGIN
```

Query Match 68.6%; Score 473.4; DB 10; Length 758;
Best Local Similarity 86.1%; Pred. No. 2.5e-120;
Matches 547; Conservative 0; Mismatches 86; Indels 2; Gaps 2;

```
Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db    115 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 174

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db    175 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 234

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db    235 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 294

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db    295 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 354

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          |||||||||||||||| |||| |||||| ||||| ||||| || ||||||||||||||||
Db    355 CAGGCCATGATGGTGACGTCCACTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 414

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db    415 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 474

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          |||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db    475 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 534

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttga-gat 479
          ||||| ||||| |||||||||||||| || ||||| |||||||||||||||||||| |||
Db    535 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGACGAT 594

Qy    480 tggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcat 539
          |||||||||||| |||||||||| || ||  |||||||  ||  |||| ||| ||||
Db    595 TGGAGAGGCTCTGTACTTGGGCATCATCTCAGCCCTGTTTTCTTTGGTAGCCGGAGTCAT 654

Qy    540 cctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagc 599
          ||| ||||||||||||||  |||||   |||||  |||||||||| |||| |||||| ||
Db    655 CCTTTGCTTTTCCTGCTCGCCCCAGGTCAATCGTACCAACTACTATGATGGCTACCAGGC 714

Qy    600 ccaacctcttgccacaaggagctctccaaggcctg 634
          || | |||||||||| | ||||||||||||  |||
Db    715 CC-AGCTCTTGCCACTAAGAGCTCTCCAAGATCTG 748

```
RESULT 12
BI148156
LOCUS       BI148156                706 bp    mRNA    linear   EST 05-JUL-2001
DEFINITION  602912333F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5053328 5',
            mRNA sequence.
ACCESSION   BI148156
VERSION     BI148156.1  GI:14608157
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11145  row: a  column: 09
            High quality sequence stop: 705.
FEATURES             Location/Qualifiers
     source          1. .706
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5053328"
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      134 a    200 c    186 g    186 t
ORIGIN
```

Query Match 66.8%; Score 460.8; DB 10; Length 706;
Best Local Similarity 84.9%; Pred. No. 7.5e-117;
Matches 574; Conservative 0; Mismatches 97; Indels 5; Gaps 5;

```
Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db     33 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTT-TGGGGCTGTTAGGC 91

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    |||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||| |
Db     92 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGT-GGTGCCAGCA-T 149

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db    150 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 209

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db    210 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 269

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          |||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db    270 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 329

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db    330 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 389

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          |||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db    390 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 449

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||| ||||| |||||||||||||| || ||||| ||||||||||||||| |||||||
Db    450 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATCTGAGATT 509

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccct-gatagctggaatcat 539
          |||||||||| |||| |||||| || ||  |||||| ||| || | |||| ||| ||||
Db    510 GGAGAGGCTCTGTACTCGGGCATCATCTCAGCCCTGTCCTCTCTCGGTAGCCGGAGTCAT 569

Qy    540 cctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagc 599
          |||  |||| |||       |||   | |||||  |||||||||| |||| |||||| ||
Db    570 CCTTCGCTTCTCCATGCTCGCCCCAGGCAATCGTACCAACTACTATGATGGCTACCAGGC 629

Qy    600 ccaacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtga 659
          ||| |||||||||||| ||||||||||||||  ||| |||| |  |||||| |||||||||
```
```

```
Db    630 CCAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGCTCAA-CAGCCAAAGCCAAGAGTGA 688

Qy    660 gttcaattcctacagc 675
          |||||| || ||||||
Db    689 GTTCAACTCATACAGC 704


RESULT 13
BG385562
LOCUS       BG385562                494 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602453858F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582214 5',
            mRNA sequence.
ACCESSION   BG385562
VERSION     BG385562.1  GI:13278350
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1304  row: c  column: 15
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4582214"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT       91 a    145 c    132 g    126 t
ORIGIN


  Query Match             66.7%;  Score 460;  DB 10;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-116;
  Matches  460;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     35 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 94

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     95 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 154

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    155 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 214

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    215 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 274

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    275 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 334

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    335 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 394

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    395 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 454

Qy    421 catgggatcctacgggacttctactcaccactggtgcctg 460
          |||||||||||||||||||||||||||||||||||||||
Db    455 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTG 494


RESULT 14
BF789255
LOCUS       BF789255                846 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  602105156F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4223461
            5', mRNA sequence.
ACCESSION   BF789255
VERSION     BF789255.1  GI:12094291
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9811  row: o  column: 14
            High quality sequence stop: 649.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4223461"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      171 a    220 c    235 g    219 t      1 others
ORIGIN


  Query Match             66.2%;  Score 456.8;  DB 10;  Length 846;
  Best Local Similarity   86.0%;  Pred. No. 1.1e-115;
  Matches  517;  Conservative    0;  Mismatches   83;  Indels    1;  Gaps    1;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||| |||||| | ||||| ||||||||||||||||||||| ||||||| || |||
Db     92 ATGGCCTCCCTTGGCGTTCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 151

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db    152 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 211
```

```
Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| ||||||||||||||||||||| ||||| ||||||||||||||||
Db    212 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 271

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db    272 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 331

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db    332 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 391

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db    392 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 451

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          |||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db    452 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGGGCTTTATCCCAGTTGCTTGGAATCTT 511

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||| ||||| |||||||||||||| || ||||| |||||||||||||| |||| ||||
Db    512 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAACTTTGCGATT 571

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          |||||||||| |||||||||| || ||   |||||| ||  || |||| ||| |||||
Db    572 GGAGAGGCTCTGTACTTGGGCATCATCTCANGCCTGTTTTCTTTGGTAGCCGGAGTCATC 631

Qy    541 ctctgcttttcctgctcatcccagag-aaatcgctccaactactacgatgcctaccaagc 599
          || |||||||||||||  ||||| | || |||  |||||||||| ||||  ||||| ||
Db    632 CTTTGCTTTTCCTGCTCGCCCCAGGGCAATTCGTACCAACTACTATGATGGTTACCAGGC 691

Qy    600 c 600
          |
Db    692 C 692
```

```
RESULT  15
AW475316
LOCUS       AW475316                750 bp    mRNA    linear   EST 24-FEB-2000
DEFINITION  un64d02.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2503395 5' similar to TR:O88552 O88552 CLAUDIN-2. ;, mRNA
            sequence.
ACCESSION   AW475316
VERSION     AW475316.1  GI:7045422
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1018247
            Seq primer: custom primer used
            High quality sequence stop: 519.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2503395"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      159 a    195 c    203 g    193 t
ORIGIN
```

```
  Query Match             65.5%;  Score 451.8;  DB 9;  Length 750;
  Best Local Similarity   85.6%;  Pred. No. 2.4e-114;
  Matches  537;  Conservative    0;  Mismatches   87;  Indels    3;  Gaps    3;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||| |||||| ||||||| ||||||||||||||||||||| ||||||| || |||
Db    127 ATGGCCTCCCTTGGCGTCCAACTGGTGGGCTACATCCTAGGCCTTTTGGGGCTGTTAGGC 186

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db    187 ACATCCATTGCCATGCTGCTTCCCAACTGGCGAACGAGTTCCTATGTTGGTGCCAGCATT 246

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| ||||||||||||||||||||| ||||| ||||||||||||||||
Db    247 GTGACGGCGGTTGGCTTTTCCAAGGGCCTCTGGATGGAGTGTGCGACACACAGCACAGGC 306

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db    307 ATCACCCAGTGCGATATCTACAGTACCCTTTTAGGACTTCCTGCTGACATCCAGGCTGCC 366

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          |||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db    367 CAGGCCATGATGGTGACGTCCAGTGCAATGTCCTCGCTGGCTTGTATTATCTCTGTGGTG 426

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db    427 GGCATGAGATGCACCGTGTTCTGCCAGGATTCTCGAGCTAAGGACAGAGTGGCTGTAGTG 486

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          |||||||||||||||||||||| ||| ||||| | || || || ||||| |||||||||
Db    487 GGTGGAGTCTTTTTCATCCTTGGTGGCATCCTGAGCTTTATCCCAGTTGCTTGGAATCTT 546

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||| ||||| |||||||||||||| || ||||| |||||||||||||| ||||||||
Db    547 CATGGCATCCTTCGGGACTTCTACTCGCCGCTGGTTCCTGACAGCATGAAAGTTGAGATT 606

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          |||||||||| |||||||||| || ||   || |||| | | |  |||| ||| |||||
Db    607 GGAGAGGCTCTGTACTTGGGCATCATCTCAGCCATGTT-TACTTTGTAGCCGGAGTCATC 665

Qy    541 ctctgcttttcctgctcatcccagagaaaatcgctccaactactacgatgcctaccaagcc 600
          || ||| | | ||||||  | ||| | |||||  |||||||||| |||| |||||| |||
Db    666 CTTTGC-TATGCTGCTCGCCGCAG-GCAATCGTACCAACTACTATGATGGCTACCAGGCC 723

Qy    601 caacctcttgccacaaggagctctcca 627
          || |||||||||| | ||||||| ||
Db    724 CAGCCTCTTGCCACTAAGAGCTCTGCA 750
```

Search completed: June 2, 2002, 20:00:07
Job time: 14617 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 18:58:39 ; Search time 108.28 Seconds
(without alignments)
1565.267 Million cell updates/sec

Title: US-09-663-600A-139_COPY_36_725
Perfect score: 690
Sequence: 1 atggcctctcttggcctcca.....acagcctgacagggtatgtg 690

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	73.8	10.7	363	2	US-08-966-316-8	Sequence 8, Appli
2	70.8	10.3	413	3	US-09-188-930-55	Sequence 55, Appl
3	43	6.2	7218	1	US-08-232-463-14	Sequence 14, Appl
4	33.8	4.9	3468	1	US-07-951-715A-4	Sequence 4, Appli
5	33.8	4.9	3468	1	US-07-951-715A-8	Sequence 8, Appli
6	33.8	4.9	3468	2	US-08-459-448A-4	Sequence 4, Appli
7	33.8	4.9	3468	2	US-08-459-448A-8	Sequence 8, Appli
8	33.8	4.9	3468	3	US-08-459-595A-4	Sequence 4, Appli
9	33.8	4.9	3468	3	US-08-459-595A-8	Sequence 8, Appli
10	33.8	4.9	3468	3	US-08-459-504B-4	Sequence 4, Appli
11	33.8	4.9	3468	3	US-08-459-504B-8	Sequence 8, Appli
12	33.8	4.9	3468	3	US-08-459-444-4	Sequence 4, Appli
13	33.8	4.9	3468	3	US-08-459-444-8	Sequence 8, Appli
14	33.8	4.9	3468	3	US-09-053-549-5	Sequence 5, Appli
15	33.8	4.9	3468	4	US-09-547-422-4	Sequence 4, Appli
16	33.8	4.9	3468	4	US-09-547-422-8	Sequence 8, Appli
17	33.8	4.9	3546	1	US-07-951-715A-10	Sequence 10, Appl
18	33.8	4.9	3546	1	US-07-951-715A-12	Sequence 12, Appl
19	33.8	4.9	3546	1	US-07-951-715A-14	Sequence 14, Appl
20	33.8	4.9	3546	1	US-07-951-715A-27	Sequence 27, Appl
21	33.8	4.9	3546	2	US-08-459-448A-10	Sequence 10, Appl
22	33.8	4.9	3546	2	US-08-459-448A-12	Sequence 12, Appl
23	33.8	4.9	3546	2	US-08-459-448A-14	Sequence 14, Appl
24	33.8	4.9	3546	2	US-08-459-448A-27	Sequence 27, Appl
25	33.8	4.9	3546	3	US-08-459-595A-10	Sequence 10, Appl
26	33.8	4.9	3546	3	US-08-459-595A-12	Sequence 12, Appl
27	33.8	4.9	3546	3	US-08-459-595A-14	Sequence 14, Appl
28	33.8	4.9	3546	3	US-08-459-595A-27	Sequence 27, Appl
29	33.8	4.9	3546	3	US-08-459-504B-10	Sequence 10, Appl
30	33.8	4.9	3546	3	US-08-459-504B-12	Sequence 12, Appl
31	33.8	4.9	3546	3	US-08-459-504B-14	Sequence 14, Appl
32	33.8	4.9	3546	3	US-08-459-504B-27	Sequence 27, Appl
33	33.8	4.9	3546	3	US-08-459-444-10	Sequence 10, Appl
34	33.8	4.9	3546	3	US-08-459-444-12	Sequence 12, Appl
35	33.8	4.9	3546	3	US-08-459-444-14	Sequence 14, Appl
36	33.8	4.9	3546	3	US-08-459-444-27	Sequence 27, Appl
37	33.8	4.9	3546	4	US-09-547-422-10	Sequence 10, Appl
38	33.8	4.9	3546	4	US-09-547-422-12	Sequence 12, Appl
39	33.8	4.9	3546	4	US-09-547-422-14	Sequence 14, Appl
40	33.8	4.9	3546	4	US-09-547-422-27	Sequence 27, Appl
41	33.8	4.9	3547	1	US-07-951-715A-16	Sequence 16, Appl
42	33.8	4.9	3547	2	US-08-459-448A-16	Sequence 16, Appl
43	33.8	4.9	3547	3	US-08-459-595A-16	Sequence 16, Appl
44	33.8	4.9	3547	3	US-08-459-504B-16	Sequence 16, Appl
45	33.8	4.9	3547	3	US-08-459-444-16	Sequence 16, Appl

ALIGNMENTS

```
RESULT 1
US-08-966-316-8
; Sequence 8, Application US/08966316
; Patent No. 5932445
; GENERAL INFORMATION:
;   APPLICANT: Lal, Preeti
;   APPLICANT: Au-Young, Janice
;   APPLICANT: Reddy, Roopa
;   APPLICANT: Murry, Lynn E.
;   APPLICANT: Mathur, Preete
;   TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Incyte Pharmaceuticals, Inc.
;     STREET: 3174 Porter Drive
;     CITY: Palo Alto
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/966,316
;     FILING DATE: Herewith
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Billings, Lucy J.
;     REGISTRATION NUMBER: 36,749
;     REFERENCE/DOCKET NUMBER: PF-0424 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 650-855-0555
;     TELEFAX: 650-845-4166
;     TELEX:
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 363 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     LIBRARY: HNT3AZT01
;     CLONE: 2417676
US-08-966-316-8
```

```
  Query Match             10.7%;  Score 73.8;  DB 2;  Length 363;
  Best Local Similarity   55.2%;  Pred. No. 4e-14;
  Matches  144;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggctttgggc 60
           |||||||||   ||  | ||  |  ||||   | ||||    || |||||     ||
Db      48 ATGGCCTCTGCCGGAATGCAGATCCTGGGAGTCGTCCTGACACTGCTGGGCTGGGTGAAT 107

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
             |||||  ||      || ||||  |||||     |   |||   ||||   ||||||
Db     108 GGCCTGGTCTCCTGTGCCCTGCCCATGTGGAAGGTGACCGCTTTCATCGGCAACAGCATC 167

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           |||   ||    |   | |   ||||||| ||||||   || |     || ||||| |||
Db     168 GTGGTGGCCCAGGTGGTGTGGGAGGGCCTGTGGATGTCCTGCGTGGTGCAGAGCACCGGC 227

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
             |  |||||  |  | ||   | | || ||||  |||||    ||| | ||||||||
Db     228 CAGATGCAGTGCAAGGTGTACGACTCACTGCTGGCGCTGCCACAGGACCTGCAGGCTGCA 287

Qy     241 caggccatgatggtgacatcc 261
           |  ||| |    || |   ||
Db     288 CGTGCCCTCTGTGTCATCGCC 308


RESULT   2
US-09-188-930-55
; Sequence 55, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 55
;   LENGTH: 413
;   TYPE: DNA
;   ORGANISM: Human
US-09-188-930-55


  Query Match             10.3%;  Score 70.8;  DB 3;  Length 413;
  Best Local Similarity   54.2%;  Pred. No. 3.9e-13;
  Matches  166;  Conservative    0;  Mismatches  137;  Indels    3;  Gaps    1;

Qy     199 tatagcacccttctgggcctgcccgctgacatccaggctgcccaggccatgatggtgaca 258
           ||   | |  | || | |||| |||| | | | |||||  | |  ||| | ||||||
Db      21 tacgactcggtgctcgccctgtccgcggccttgcaggccactcgagccctaatggtggtc 80

Qy     259 tccagtgcaatctcctccctggcctgcattatctctgtggtgggcatgagatgcacagtc 318
           |||   |   |   || |||||||    || |  |   | |||||||||  |||||   |
Db      81 tccctggtgctgggcttcctggccatgtttgtggccacgatgggcatgaagtgcacgcgc 140

Qy     319 ttctgccaggaatcccgagccaaaga---cagagtggcggtagcaggtggagtctttttc 375
           |   |    ||   |  ||  ||  |   | |  | ||  | |  || ||   |  |||||
Db     141 tgtgggggagacgacaaagtgaagaaggcccgtatagccatgggtggaggcataatttttc 200

Qy     376 atccttggaggcctcctgggattcattcctgttgcctggaatcttcatgggatcctacgg 435
           ||| | | ||| ||    |  ||  |  ||    ||||| ||   |||  |||  |
Db     201 atcgtggcaggtcttgccgccttggtagcttgctcctggtatggccatcagattgtcaca 260

Qy     436 gacttctactcaccactggtgcctgacagcatgaaatttgagattggagaggctctttac 495
           |||||  ||    ||  || | |||  || ||| || | |||| ||||    ||  | |
Db     261 gacttttataacccttgatccctaccaacattaagtatgagtttggccctgccatcttt 320

Qy     496 ttgggc 501
             | |||
Db     321 attggc 326


RESULT   3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             6.2%;  Score 43;  DB 1;  Length 7218;
  Best Local Similarity   6.0%;  Pred. No. 0.0014;
  Matches   19;  Conservative  169;  Mismatches  129;  Indels    0;  Gaps    0;

Qy     329 aatcccgagccaaagacagagtggcggtagcaggtggagtctttttcatccttggaggcc 388
           |||  ||||||          || | || |||  | |   :::::::: :::::     ::
Db    1029 AATTCCGAGCTTGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYY 1088

Qy     389 tcctgggattcattcctgttgcctggaatcttcatgggatcctacgggacttctactcac 448
           ::::    ::: ::::: :: :::    ::::: :    :::: :    ::::: ::: :
Db    1089 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1148

Qy     449 cactggtgcctgacagcatgaaatttgagattggagaggctctttacttgggcattattt 508
           : ::  : :::  :  : :    :::    ::       :::::: :::   : :: :::
Db    1149 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1208

Qy     509 cttccctgttctccctgatagctggaatcatcctctgcttttcctgctcatcccagagaa 568
```

```
               ::::::: ::::::::  :  ::     :: :::::: :::::::: ::: ::::
Db    1209 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1268

Qy     569 atcgctccaactactacgatgcctaccaagcccaacctcttgccacaaggagctctccaa 628
              :: ::::  :: :: :  : ::: ::   :::  :::::: :: :      ::::::
Db    1269 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1328

Qy     629 ggcctggtcaacctccc 645
             :::  ::  ::::::
Db    1329 YYYYYYYYYYYYYYYYY 1345
```

```
RESULT   4
US-07-951-715A-4
; Sequence 4, Application US/07951715A
; Patent No. 5625136
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/951,715A
;      FILING DATE:  25-SEP-1992
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  S-18805/A/CGC 1577/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8615
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..3468
;      OTHER INFORMATION:  /product= "Full length synthetic
;      OTHER INFORMATION:  maize optimized"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 3 as synful.mod.  This sequenc
;      OTHER INFORMATION:  identical to flsynbt.fin as disclosed in Figure 1."
US-07-951-715A-4

  Query Match              4.9%;  Score 33.8;  DB 1;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |           |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |      |||||| ||   |||  |     ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
            || ||||| | | |   | |    | | |||  ||| ||| || |       | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
            | | ||||
Db    1583 GCTACGCCA 1591
```

```
RESULT   5
US-07-951-715A-8
; Sequence 8, Application US/07951715A
; Patent No. 5625136
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/951,715A
;      FILING DATE:  25-SEP-1992
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  S-18805/A/CGC 1577/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8615
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3465
;      OTHER INFORMATION:  /product= "Full-length, hybrid,
;      OTHER INFORMATION:  partially maize optimized cryIA(b)"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 7 as contained in pCIB4434."
US-07-951-715A-8


  Query Match             4.9%;  Score 33.8;  DB 1;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |          |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     |||||| ||   |||  |    ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
           || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
           | | ||||
Db    1583 GCTACGCCA 1591


RESULT   6
US-08-459-448A-4
; Sequence 4, Application US/08459448A
; Patent No. 5859336
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5859336artis Corporation
;      STREET:  Patent & Trademark Dept., 520 White Plains
;      STREET:  Rd., POB 2005
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-9005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,448A
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/951,715
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40403
;      REFERENCE/DOCKET NUMBER:  CGC 1577/CIP/DIV4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8582
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..3468
;      OTHER INFORMATION:  /product= "Full length synthetic
;      OTHER INFORMATION:  maize optimized"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 3 as synful.mod.  This sequenc
;      OTHER INFORMATION:  identical to flsynbt.fin as disclosed in Figure 1."
US-08-459-448A-4


  Query Match             4.9%;  Score 33.8;  DB 2;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |          |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     |||||| ||   |||  |    ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
           || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
           | | ||||
Db    1583 GCTACGCCA 1591


RESULT   7
US-08-459-448A-8
; Sequence 8, Application US/08459448A
; Patent No. 5859336
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
```

```
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5859336artis Corporation
;      STREET:  Patent & Trademark Dept., 520 White Plains
;      STREET:  Rd., POB 2005
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-9005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,448A
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/951,715
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40403
;      REFERENCE/DOCKET NUMBER:  CGC 1577/CIP/DIV4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8582
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3465
;      OTHER INFORMATION:  /product= "Full-length, hybrid,
;      OTHER INFORMATION:  partially maize optimized cryIA(b)"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 7 as contained in pCIB4434."
US-08-459-448A-8


  Query Match             4.9%;  Score 33.8;  DB 2;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |          |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     |||||| ||   |||  |   ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
            || ||||| | | |  | |   | | |||  ||| ||| || | |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
           | | ||||
Db    1583 GCTACGCCA 1591


RESULT   8
US-08-459-595A-4
; Sequence 4, Application US/08459595A
; Patent No. 6018104
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6018104artis Corporation
;      STREET:  Patent & Trademark Dept., 520 White Plains
;      STREET:  Rd., POB 2005
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-9005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,595A
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/951,715
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40403
;      REFERENCE/DOCKET NUMBER:  CGC 1577/CIP/DIV3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8582
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..3468
;      OTHER INFORMATION:  /product= "Full length synthetic
;      OTHER INFORMATION:  maize optimized"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 3 as synful.mod.  This sequence i
;      OTHER INFORMATION:  identical to flsynbt.fin as disclosed in Figure 1."
US-08-459-595A-4


  Query Match             4.9%;  Score 33.8;  DB 3;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |          |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     ||||| ||  |||  |    ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
           || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
           | | ||||
Db    1583 GCTACGCCA 1591


RESULT   9
US-08-459-595A-8
; Sequence 8, Application US/08459595A
; Patent No. 6018104
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6018104artis Corporation
;      STREET:  Patent & Trademark Dept., 520 White Plains
;      STREET:  Rd., POB 2005
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-9005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,595A
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/951,715
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40403
;      REFERENCE/DOCKET NUMBER:  CGC 1577/CIP/DIV3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8582
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3465
;      OTHER INFORMATION:  /product= "Full-length, hybrid,
;      OTHER INFORMATION:  partially maize optimized cryIA(b)"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 7 as contained in pCIB4434."
US-08-459-595A-8


  Query Match             4.9%;  Score 33.8;  DB 3;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |          |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     ||||| ||  |||  |    ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
           || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
           | | ||||
Db    1583 GCTACGCCA 1591


RESULT  10
US-08-459-504B-4
; Sequence 4, Application US/08459504B
; Patent No. 6075185
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
```

```
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6075185artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,504B
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/459,595
;      FILING DATE:  02-JUN-1995
;      APPLICATION NUMBER:  US 07/951,715
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1577/CIP/DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8587
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..3468
;      OTHER INFORMATION:  /product= "Full length synthetic
;      OTHER INFORMATION:  maize optimized"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 3 as synful.mod.  This sequence i
;      OTHER INFORMATION:  identical to flsynbt.fin as disclosed in Figure 1."
US-08-459-504B-4


  Query Match             4.9%;  Score 33.8;  DB 3;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || | |         |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     |||||| ||   |||  |    ||     |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
            || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
            | | ||||
Db    1583 GCTACGCCA 1591


RESULT  11
US-08-459-504B-8
; Sequence 8, Application US/08459504B
; Patent No. 6075185
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6075185artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,504B
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/459,595
;      FILING DATE:  02-JUN-1995
;      APPLICATION NUMBER:  US 07/951,715
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1577/CIP/DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8587
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3465
;      OTHER INFORMATION:  /product= "Full-length, hybrid,
;      OTHER INFORMATION:  partially maize optimized cryIA(b)"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 7 as contained in pCIB4434."
```

```
US-08-459-504B-8


  Query Match             4.9%;  Score 33.8;  DB 3;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |         |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     |||||| ||   |||  |    ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
            || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
           | | ||||
Db    1583 GCTACGCCA 1591


RESULT  12
US-08-459-444-4
; Sequence 4, Application US/08459444A
; Patent No. 6121014
;   GENERAL INFORMATION:
;        APPLICANT: Koziel, Michael G.
;                   Desai, Nalini M.
;                   Lewis, Kelly S.
;                   Kramer, Vance C.
;                   Warren, Gregory W.
;                   Evola, Stephen V.
;                   Crossland, Lyle D.
;                   Wright, Martha S.
;                   Merlin, Ellis J.
;                   Launis, Karen L.
;        TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED
;                            NUCLEIC ACID CODING SEQUENCE
;        NUMBER OF SEQUENCES: 94
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: No. 6121014artis Agribusiness Biotechnology Research, Inc.
;             STREET: 3054 Cornwallis Road
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/459,444A
;             FILING DATE: 02-Jun-1995
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 07/951,715
;             FILING DATE: 25-SEP-1992
;             APPLICATION NUMBER: US 07/772,027
;             FILING DATE: 04-OCT-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Meigs, J. Timothy
;             REGISTRATION NUMBER: 38,241
;             REFERENCE/DOCKET NUMBER: S-18805/P1/CGC1577/CIP/DIV6
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (919)541-8587
;             TELEFAX: (919)541-8689
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 3468 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;             DESCRIPTION: /desc = "Synthetic DNA"
;        HYPOTHETICAL: NO
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION:  1..3468
;             OTHER INFORMATION:  /product= "Full length synthetic
;             maize optimized"
;             /note= "Disclosed in Figure 3 as synful.mod.  This sequence is
;             identical to flsynbt.fin as disclosed in Figure 1."
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-459-444-4


  Query Match             4.9%;  Score 33.8;  DB 3;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |         |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     |||||| ||   |||  |    ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
            || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
           | | ||||
Db    1583 GCTACGCCA 1591


RESULT  13
US-08-459-444-8
; Sequence 8, Application US/08459444A
; Patent No. 6121014
;   GENERAL INFORMATION:
;        APPLICANT: Koziel, Michael G.
;                   Desai, Nalini M.
;                   Lewis, Kelly S.
;                   Kramer, Vance C.
;                   Warren, Gregory W.
;                   Evola, Stephen V.
;                   Crossland, Lyle D.
;                   Wright, Martha S.
;                   Merlin, Ellis J.
;                   Launis, Karen L.
;        TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED
;                            NUCLEIC ACID CODING SEQUENCE
;        NUMBER OF SEQUENCES: 94
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: No. 6121014artis Agribusiness Biotechnology Research, Inc.
;             STREET: 3054 Cornwallis Road
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/459,444A
;             FILING DATE: 02-Jun-1995
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
```

```
;            APPLICATION NUMBER: US 07/951,715
;            FILING DATE: 25-SEP-1992
;            APPLICATION NUMBER: US 07/772,027
;            FILING DATE: 04-OCT-1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Meigs, J. Timothy
;            REGISTRATION NUMBER: 38,241
;            REFERENCE/DOCKET NUMBER: S-18805/P1/CGC1577/CIP/DIV6
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (919)541-8587
;            TELEFAX: (919)541-8689
;  INFORMATION FOR SEQ ID NO: 8:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 3468 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;            DESCRIPTION: /desc = "Synthetic DNA"
;       HYPOTHETICAL: NO
;       FEATURE:
;            NAME/KEY: CDS
;            LOCATION: 1..3465
;            OTHER INFORMATION: /product= "Full-length, hybrid,
;            partially maize optimized cryIA(b)"
;            /note= "Disclosed in Figure 7 as contained in pCIB4434."
;       SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-459-444-8


  Query Match             4.9%;  Score 33.8;  DB 3;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |          |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     |||||| ||   |||  |    ||     |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
            || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
            | | ||||
Db    1583 GCTACGCCA 1591


RESULT  14
US-09-053-549-5
; Sequence 5, Application US/09053549
; Patent No. 6121521
;  GENERAL INFORMATION:
;    APPLICANT:  Desai, Nalini
;    TITLE OF INVENTION:  No. 6121521el Insecticidal Protein and Gene
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6121521artis Corporation
;      STREET:  3054 Cornwallis Rd.
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/053,549
;      FILING DATE:  01-APR-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1995
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..3468
;      OTHER INFORMATION:  /product= "Full length synthetic
;      OTHER INFORMATION:  maize optimized"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3468
US-09-053-549-5


  Query Match             4.9%;  Score 33.8;  DB 3;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |          |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     |||||| ||   |||  |    ||     |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
            || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
            | | ||||
Db    1583 GCTACGCCA 1591


RESULT  15
US-09-547-422-4
; Sequence 4, Application US/09547422
; Patent No. 6320100
;  GENERAL INFORMATION:
;       APPLICANT: Koziel, Michael G.
;                  Desai, Nalini M.
;                  Lewis, Kelly S.
;                  Kramer, Vance C.
;                  Warren, Gregory W.
;                  Evola, Stephen V.
;                  Crossland, Lyle D.
;                  Wright, Martha S.
;                  Merlin, Ellis J.
;                  Launis, Karen L.
;       TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;                           INSECTICIDAL ACTIVITY IN MAIZE
;       NUMBER OF SEQUENCES: 94
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: No. 6320100artis Agribusiness Biotechnology Research, Inc.
;            STREET: 3054 Cornwallis Road
;            CITY: Research Triangle Park
;            STATE: NC
```

```
;           COUNTRY: USA
;           ZIP: 27709
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/547,422
;           FILING DATE: 11-Apr-2000
;           CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 08/459,595
;           FILING DATE: 02-JUN-1995
;           APPLICATION NUMBER: US 07/951,715
;           FILING DATE: 25-SEP-1992
;           APPLICATION NUMBER: US 07/772,027
;           FILING DATE: 04-OCT-1991
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Meigs, J. Timothy
;           REGISTRATION NUMBER: 38,241
;           REFERENCE/DOCKET NUMBER: S-18805H
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (919)541-8587
;           TELEFAX: (919)541-8689
;   INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 3468 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;           DESCRIPTION: /desc = "Synthetic DNA"
;       HYPOTHETICAL: NO
;       FEATURE:
;           NAME/KEY:  misc_feature
;           LOCATION:  1..3468
;           OTHER INFORMATION:  /product= "Full length synthetic
;           maize optimized"
;           /note= "Disclosed in Figure 3 as synful.mod.  This sequence is
;           identical to flsynbt.fin as disclosed in Figure 1."
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-547-422-4


  Query Match             4.9%;  Score 33.8;  DB 4;  Length 3468;
  Best Local Similarity   48.7%;  Pred. No. 0.83;
  Matches   92;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      59 gcacactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagca 118
           |||  | | |  ||  | | | ||  | |  ||  || |          |||  |||||
Db    1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACCAACCTGGGCAGCGGCACCAGCG 1462

Qy     119 ttgtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacag 178
           | ||||  |     ||||||  ||   |||  |    ||    |  |||  | | ||
Db    1463 TGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAGCCCCGGCCAGA 1522

Qy     179 gcatcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctg 238
           || ||||| | | |  | |   | | |||  ||| ||| || |     | |||
Db    1523 TCAGCACCCTGCGCGTGAACATCACCGCCCCCCTGAGCCAGCGCTACCGCGTCCGCATCC 1582

Qy     239 cccaggcca 247
           | | ||||
Db    1583 GCTACGCCA 1591


Search completed: June  2, 2002, 18:58:54
Job time: 11079 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 20:08:48 ; Search time 523.16 Seconds
(without alignments)
2264.452 Million cell updates/sec

Title: US-09-663-600A-139_COPY_36_725
Perfect score: 690
Sequence: 1 atggcctctcttggcctcca.....acagcctgacagggtatgtg 690

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	690	100.0	1400	20	AAX97865	Human secreted pro
2	690	100.0	1475	21	AAA37060	Human PRO1356 (UNQ
3	690	100.0	1475	22	AAS15360	cDNA encoding huma
4	690	100.0	1475	22	AAS21489	Human cDNA sequenc
5	690	100.0	1475	22	AAF92097	Human PRO1356 cDNA
6	690	100.0	1475	22	AAF54296	DNA encoding prote
7	690	100.0	1481	22	AAD08497	Human secreted pro
8	690	100.0	1918	24	ABA04424	Human SP82 protein
9	690	100.0	2742	21	AAA12585	cDNA encoding a me
10	690	100.0	2863	22	AAF93769	Human cDNA encodin
11	689.6	99.9	1441	24	AAS16180	Human CASB81 polyn
12	689.2	99.9	1524	20	AAX97818	Human secreted pro
13	688.4	99.8	1540	22	AAI58013	Human polynucleoti
c 14	688.4	99.8	1761	22	AAI59799	Human polynucleoti
15	546	79.1	779	24	AAS16181	Murine CASB81 poly
16	546	79.1	791	21	AAZ89137	Murine clodin 2 cD
17	417.4	60.5	530	22	AAD08518	Human secreted pro
18	388.8	56.3	467	21	AAC98817	Human pancreatic c
19	361.6	52.4	615	22	AAF93963	Primer specific fo
20	339.2	49.2	405	20	AAX41059	Human secreted pro
21	263	38.1	265	22	AAH34986	Human colon cancer
c 22	251.8	36.5	1046	22	AAD08519	Human secreted pro
23	215.8	31.3	1174	21	AAA37113	Human PRO1571 (UNQ
24	215.8	31.3	1174	22	AAS46101	Human DNA encoding
25	215.8	31.3	1174	22	AAF54431	Probe #47 used in
26	196.6	28.5	1361	22	AAK51692	Human polynucleoti
27	193.4	28.0	1381	22	AAK52676	Human polynucleoti
28	165	23.9	182	22	AAF67470	Novel human polynu
29	165	23.9	324	20	AAX41060	Human secreted pro
30	154.8	22.4	1779	20	AAV80582	Kidney injury asso
31	146.8	21.3	766	23	AAS92167	DNA encoding novel
32	143.8	20.8	300	20	AAZ13117	Human gene express
33	141	20.4	681	22	AAF60261	Human Claudin-7 cD
34	141	20.4	1014	21	AAC95568	Human secreted pro
35	141	20.4	1253	20	AAZ42051	Human endometrium
36	141	20.4	1546	21	AAZ47929	Human apoptosis as
37	141	20.4	1583	22	AAH34904	Human colon cancer
38	141	20.4	1711	19	AAV43617	Human secreted pro
39	141	20.4	1787	22	AAF98719	Human late stage o
40	139.4	20.2	1705	20	AAV84506	Human secreted pro
41	139.4	20.2	1705	22	ABA83289	Human secreted pro
42	138.8	20.1	1665	21	AAZ36702	Nucleotide sequenc
43	138.8	20.1	1665	22	ABA83103	Claudin 4 ovarian
44	138.8	20.1	1717	22	AAF98710	Human late stage o
45	138.8	20.1	1776	22	AAH34872	Human colon cancer

ALIGNMENTS

RESULT 1
AAX97865
ID AAX97865 standard; cDNA; 1400 BP.
XX
AC AAX97865;
XX
DT 23-SEP-1999 (first entry)
XX
DE Human secreted protein encoding cDNA #53.
XX
KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW diagnostic; gene therapy; chromosome mapping; secretion vector; ss.
XX
OS Homo sapiens.
XX
PN WO9925825-A2.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-IB01862.
XX
PR 04-SEP-1998; 98US-0099273.
PR 13-NOV-1997; 97US-0066677.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
PR 10-AUG-1998; 98US-0096116.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX

```
DR   WPI; 1999-347472/29.
DR   P-PSDB; AAY36181.
XX
PT   Extended cDNAs encoding secreted proteins
XX
PS   Claim 1; Page 254-255; 307pp; English.
XX
CC   AAX97813-X97906 represent extended cDNA's which encode novel human
CC   secreted proteins (see AAY36129-Y36222) and which have cytostatic,
CC   thrombotic and osteopathic activity. The extended cDNAs can be used to
CC   express secreted proteins or parts of them or to obtain antibodies
CC   capable of binding to the secreted proteins. They may also be used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   Uses also include design of expression vectors and secretion vectors.
XX
SQ   Sequence 1400 BP; 326 A; 392 C; 346 G; 336 T; 0 other;


  Query Match           100.0%;  Score 690;  DB 20;  Length 1400;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-204;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     36 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 95

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     96 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 155

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 215

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 275

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    276 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 335

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    336 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 395

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    396 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 455

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    456 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 515

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    516 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 575

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    576 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 635

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    636 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 695

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    696 ttcaattcctacagcctgacagggtatgtg 725


RESULT   2
AAA37060

ID   AAA37060 standard; cDNA; 1475 BP.
XX
AC   AAA37060;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1356 (UNQ705) cDNA sequence SEQ ID NO:133.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20111.
XX
PR   01-SEP-1998;   98US-0098716.
PR   01-SEP-1998;   98US-0098749.
PR   01-SEP-1998;   98US-0098750.
PR   02-SEP-1998;   98US-0098803.
PR   02-SEP-1998;   98US-0098821.
PR   02-SEP-1998;   98US-0098843.
PR   09-SEP-1998;   98US-0099536.
PR   09-SEP-1998;   98US-0099596.
PR   09-SEP-1998;   98US-0099598.
PR   09-SEP-1998;   98US-0099602.
PR   09-SEP-1998;   98US-0099642.
PR   10-SEP-1998;   98US-0099741.
PR   10-SEP-1998;   98US-0099754.
PR   10-SEP-1998;   98US-0099763.
PR   10-SEP-1998;   98US-0099792.
PR   10-SEP-1998;   98US-0099808.
PR   10-SEP-1998;   98US-0099812.
PR   10-SEP-1998;   98US-0099815.
PR   10-SEP-1998;   98US-0099816.
PR   15-SEP-1998;   98US-0100385.
PR   15-SEP-1998;   98US-0100388.
PR   15-SEP-1998;   98US-0100390.
PR   16-SEP-1998;   98US-0100584.
PR   16-SEP-1998;   98US-0100627.
PR   16-SEP-1998;   98US-0100661.
PR   16-SEP-1998;   98US-0100662.
PR   16-SEP-1998;   98US-0100664.
PR   17-SEP-1998;   98US-0100683.
PR   17-SEP-1998;   98US-0100684.
PR   17-SEP-1998;   98US-0100710.
PR   17-SEP-1998;   98US-0100711.
PR   17-SEP-1998;   98US-0100919.
PR   17-SEP-1998;   98US-0100930.
PR   18-SEP-1998;   98US-0100848.
PR   18-SEP-1998;   98US-0100849.
PR   18-SEP-1998;   98US-0101014.
PR   18-SEP-1998;   98US-0101068.
PR   18-SEP-1998;   98US-0101071.
PR   22-SEP-1998;   98US-0101279.
PR   23-SEP-1998;   98US-0101471.
PR   23-SEP-1998;   98US-0101472.
PR   23-SEP-1998;   98US-0101474.
PR   23-SEP-1998;   98US-0101475.
PR   23-SEP-1998;   98US-0101476.
PR   23-SEP-1998;   98US-0101477.
PR   23-SEP-1998;   98US-0101479.
PR   24-SEP-1998;   98US-0101738.
PR   24-SEP-1998;   98US-0101741.
PR   24-SEP-1998;   98US-0101743.
PR   24-SEP-1998;   98US-0101915.
PR   24-SEP-1998;   98US-0101916.
PR   29-SEP-1998;   98US-0102207.
PR   29-SEP-1998;   98US-0102240.
```

PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

XX
DR WPI; 2000-237871/20.
DR P-PSDB; AAY99378.
XX
PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS Claim 2; Fig 77; 773pp; English.
XX
CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;

Query Match 100.0%; Score 690; DB 21; Length 1475;
Best Local Similarity 100.0%; Pred. No. 7.6e-204;
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
Db 122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy 61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
Db 182 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 241

Qy 121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
Db 242 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 301

Qy 181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
Db 302 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 361

Qy 241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
Db 362 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 421

Qy 301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
Db 422 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 481

Qy 361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
Db 482 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 541

Qy 421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
Db 542 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 601

Qy 481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
Db 602 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 661

Qy 541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
Db 662 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 721

Qy 601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
Db 722 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 781

Qy 661 ttcaattcctacagcctgacagggtatgtg 690
||||||||||||||||||||||||||||||

```
Db    782 ttcaattcctacagcctgacagggtatgtg 811


RESULT   3
AAS15360
ID   AAS15360 standard; cDNA; 1475 BP.
XX
AC   AAS15360;
XX
DT   16-JAN-2002  (first entry)
XX
DE   cDNA encoding human PRO1356 polypeptide.
XX
KW   Human; PRO1356; clone DNA64886-1601; immune-related disorder;
KW   inflammatory disorder; infectious disorder; immunodeficiency disorder;
KW   autoimmune disorder; renal disease; demyelinating disease; skin disease;
KW   neoplasia; transplantation associated disease; gene therapy;
KW   immunosuppressive; anti-inflammatory; antiasthmatic; antidiabetic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             122..814
FT                   /*tag= a
FT                   /product= "PRO1356 polypeptide"
FT   sig_peptide     122..193
FT                   /*tag= b
FT   mat_peptide     194..811
FT                   /*tag= c
XX
PN   WO200166740-A2.
XX
PD   13-SEP-2001.
XX
PF   01-MAR-2001; 2001WO-US06666.
XX
PR   03-MAR-2000; 2000US-187202P.
PR   21-MAR-2000; 2000US-191015P.
PR   30-MAY-2000; 2000WO-US14941.
PR   05-JUN-2000; 2000US-209832P.
PR   24-AUG-2000; 2000WO-US23328.
PR   01-DEC-2000; 2000WO-US32678.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Fong S,  Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL;
PI   Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-625876/72.
DR   P-PSDB; AAU09178.
XX
PT   Nucleic acids encoding PRO polypeptides, useful for detecting and
PT   treating immune related diseases and disorders in mammals including
PT   autoimmune diseases, inflammatory diseases and asthma -
XX
PS   Claim 2; Fig 1; 122pp; English.
XX
CC   The present invention relates to the isolation of 9 novel human PRO
CC   polypeptides (AAU09178-AAU09186) and the cDNA sequences encoding them.
CC   The novel PRO polypeptides include PRO1356, PRO1268, PRO1884, PRO3444,
CC   PRO3151, PRO4322, PRO9964, PRO10008 and PRO19598. The cDNA sequences
CC   encoding these PRO polypeptides have been designated as clones
CC   DNA64886-1601, DNA64903-1553, DNA84318-2520, DNA87997, DNA89273,
CC   DNA92223-2567, DNA96973, DNA101921 and DNA145887 respectively.
CC   Compositions (e.g. vaccines) containing PRO polypeptides and methods of
CC   using these compositions are useful in the treatment and diagnosis of
CC   immune-related disorders. Such disorders include immune-mediated
CC   inflammatory disorders (e.g. osteoarthritis), non-immune-mediated
CC   inflammatory  disorders (e.g. diabetes mellitus), infectious disorders
CC   (e.g. granulomatous hepatitis), immunodeficiency disorders (e.g. AIDS),
CC   autoimmune disorders (e.g. rheumatoid arthritis), immune-related renal
CC   diseases (e.g. cirrhosis), demyelinating diseases of the peripheral or
CC   central nervous system (e.g. Guillain-Barre syndrome), immune-mediated
CC   skin diseases (e.g. contact dermatitis), neoplasias and transplantation
CC   associated diseases. The polynucleotide sequences of the invention may
CC   be used in gene therapy. AAS15360-AAS15368 represent cDNA sequences
CC   encoding for the novel human PRO polypeptides of the invention.
XX
SQ   Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;


  Query Match           100.0%;  Score 690;  DB 22;  Length 1475;
  Best Local Similarity  100.0%;  Pred. No. 7.6e-204;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 241

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 301

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 361

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 421

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 481

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 541

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 601

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 661

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 721

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 781

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    782 ttcaattcctacagcctgacagggtatgtg 811


RESULT   4
AAS21489
ID   AAS21489 standard; cDNA; 1475 BP.
XX
AC   AAS21489;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human cDNA sequence encoding for PRO1356 polypeptide.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
```

```
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   09-DEC-1999;   99US-0170262.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Beresini M,  Deforge L,  Desnoyers L,  Filvaroff E,  Gao W;
PI   Gerritsen ME,  Goddard A,  Godowski PJ,  Gurney AL,  Sherwood S;
PI   Smith V,  Stewart TA,  Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-408281/43.
DR   P-PSDB; AAU12417.
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical  -
XX
PS   Claim 3; Fig 491; 813pp; English.
XX
CC   AAS21244-AAS21518 encode for novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;


  Query Match             100.0%;  Score 690;  DB 22;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-204;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 241

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 301

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 361

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 421

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 481

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 541

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 601

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 661

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 721

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 781

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db     782 ttcaattcctacagcctgacagggtatgtg 811


RESULT   5
AAF92097
ID   AAF92097 standard; cDNA; 1475 BP.
XX
AC   AAF92097;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human PRO1356 cDNA.
XX
KW   Human; PRO protein; mapping; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200116318-A2.
XX
PD   08-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-US23328.
XX
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   07-DEC-1999;   99US-0169495.
PR   09-DEC-1999;   99US-0170262.
PR   11-JAN-2000; 2000US-0175481.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   01-MAR-2000; 2000WO-US05601.
PR   03-MAR-2000; 2000US-0187202.
PR   25-APR-2000; 2000US-0199397.
PR   22-MAY-2000; 2000WO-US14042.
PR   05-JUN-2000; 2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   P-PSDB; AAB87565.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 2; Fig 79; 278pp; English.
XX
CC   The present sequence is the coding sequence for a human PRO polypeptide
CC   (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
CC   antagonists or anti-PRO antibodies are useful for preparation of a
CC   medicament useful in the treatment of a condition which is responsive to
CC   the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
CC   protein may also be employed as molecular weight markers for protein
CC   electrophoresis. The PRO coding sequence has applications in molecular
CC   biology, including use as hybridisation probes, and in chromosome and
CC   gene mapping.
XX
SQ   Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;


  Query Match           100.0%;  Score 690;  DB 22;  Length 1475;
  Best Local Similarity  100.0%;  Pred. No. 7.6e-204;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 241

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 301

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 361

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 421

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 481

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 541

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 601

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 661

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 721

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 781

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    782 ttcaattcctacagcctgacagggtatgtg 811


RESULT   6
AAF54296
ID   AAF54296 standard; DNA; 1475 BP.
XX
AC   AAF54296;
XX
DT   02-APR-2001  (first entry)
XX
DE   DNA encoding protein of the invention #39.
XX
KW   Secreted; transmembrane; gene therapy; ss.
XX
OS   Unidentified.
XX
PN   WO200078961-A1.
XX
PD   28-DEC-2000.
XX
PF   18-FEB-2000; 2000WO-US04342.
XX
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   16-DEC-1999;   99WO-US30095.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Botstein D,  Desnoyers L,  Eaton DL,  Ferrara N,  Fong S;
PI   Gao W,  Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Pan J,  Paoni NF,  Roy MA,  Smith V,  Stewart TA,  Tumas D;
PI   Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2001-071395/08.
XX
PT   Secreted and transmembrane proteins and nucleic acids designated PRO,
PT   useful as hybridization probes, in chromosome and gene mapping and gene
PT   therapy -
XX
PS   Claim 2; Fig 77; 787pp; English.
XX
```

```
CC   The present invention relates to secreted and transmembrane proteins.
CC   These proteins and the DNA encoding them may be used as hybridization
CC   probes, in chromosome and gene mapping and in the generation of
CC   anti-sense RNA and DNA. They may also be used used to generate either
CC   transgenic animals or knockout animals which are in turn useful for
CC   development and screening of therapeutically useful reagents.
CC   The nucleic acids may also be used in gene therapy.
XX
SQ   Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;


  Query Match             100.0%;  Score 690;  DB 22;  Length 1475;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-204;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 181

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 241

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 301

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 361

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 421

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 481

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 541

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 601

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 661

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 721

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 781

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
           |||||||||||||||||||||||||||||
Db     782 ttcaattcctacagcctgacagggtatgtg 811


RESULT   7
AAD08497
ID   AAD08497 standard; cDNA; 1481 BP.
XX
AC   AAD08497;
XX
DT   09-AUG-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 10 cDNA clone HTPIH83, SEQ ID NO:20.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             118..810
FT                   /*tag= a
FT                   /product= "Human secreted protein precursor"
FT   sig_peptide     118..189
FT                   /*tag= b
FT   mat_peptide     190..807
FT                   /*tag= c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200136432-A2.
XX
PD   25-MAY-2001.
XX
PF   15-NOV-2000; 2000WO-US31162.
XX
PR   19-NOV-1999;   99US-0166415.
PR   30-JUN-2000; 2000US-0215136.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Baker KP,  Young PE;
XX
DR   WPI; 2001-343793/36.
DR   P-PSDB; AAE04207.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; Page 410; 509pp; English.
XX
CC   AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC   protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC   AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   18 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin ageing due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
```

```
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 1481 BP; 353 A; 405 C; 375 G; 348 T; 0 other;


  Query Match             100.0%;  Score 690;  DB 22;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-204;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 177

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 237

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     238 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 297

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 357

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 417

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     418 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 477

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     478 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 537

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     538 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 597

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     598 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 657

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     658 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 717

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     718 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 777

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db     778 ttcaattcctacagcctgacagggtatgtg 807


RESULT   8
ABA04424
ID   ABA04424 standard; cDNA; 1918 BP.
XX
AC   ABA04424;
XX
DT   11-MAR-2002  (first entry)
XX
DE   Human SP82 protein encoding cDNA SEQ ID NO:1/3.
XX
KW   Human; SP82; cancer suppression; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             520..1212
FT                   /*tag= a
FT                   /product= "SP82"
XX
PN   CN1313315-A.
XX
PD   19-SEP-2001.
XX
PF   13-MAR-2000; 2000CN-0111989.
XX
PR   13-MAR-2000; 2000CN-0111989.
XX
PA   (SHAN-) SHANGHAI INST ONCOLOGY.
XX
PI   Gu J,  Yang S;
XX
DR   WPI; 2002-042193/06.
DR   P-PSDB; ABB04707.
XX
PT   New human protein able to suppress growth of cancer cells and its
PT   encoding polynucleotide sequence -
XX
PS   Claim 5; Page 11-12 (Disclosure); 42pp; Chinese.
XX
CC   The present sequence encodes human SP82 protein, which has cancer-
CC   suppressing activity. The present invention also describes a method
CC   for the preparation of the protein by recombination, and the application
CC   of the protein in treating diseases such as cancer.
XX
SQ   Sequence 1918 BP; 486 A; 525 C; 473 G; 434 T; 0 other;


  Query Match             100.0%;  Score 690;  DB 24;  Length 1918;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-204;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 579

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 639

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 699

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 759

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 819

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 879

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 939

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 999

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 1059
```

```
Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 1119

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 1179

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
           |||||||||||||||||||||||||||||
Db    1180 ttcaattcctacagcctgacagggtatgtg 1209
```

RESULT 9
AAA12585
ID AAA12585 standard; cDNA; 2742 BP.
XX
AC AAA12585;
XX
DT 25-JUL-2000 (first entry)
XX
DE cDNA encoding a membrane associated organizational protein (HJNCT).
XX
KW Human; membrane associated organizational protein; HJNCT;
KW cell proliferative disorder; cancer; autoimmune disorder;
KW inflammatory disorder; neurological disorder; developmental disorder;
KW vesicle trafficking; reproductive disorder; gastrointestinal disorder;
KW renal disorder; atherosclerosis; leukaemia; rheumatoid arthritis;
KW Alzheimer's disease; anxiety; diabetes; ovulatory defect; renal failure;
KW irritable bowel syndrome; allergy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 119..811
FT /*tag= a
FT /product= "membrane associated organizational protein"
XX
PN WO200018915-A2.
XX
PD 06-APR-2000.
XX
PF 23-SEP-1999; 99WO-US22082.
XX
PR 25-SEP-1998; 98US-0155215.
PR 13-OCT-1998; 98US-0155251.
PR 04-MAY-1999; 99US-0172228.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR, Lu AD, Tang YT;
XX
DR WPI; 2000-293154/25.
DR P-PSDB; AAY84609.
XX
PT Human membrane associated organizational protein and nucleic acid
PT sequences useful in the diagnosis, treatment and prevention of cell
PT proliferative associated disorders e.g. cancer, rheumatoid arthritis
PT and Alzheimer's disease -
XX
PS Claim 9; Page 79-80; 84pp; English.
XX
CC The present sequence encodes a human membrane associated organizational
CC protein (HJNCT). HJNCT is used for the diagnosis, treatment and
CC prevention of cell proliferative disorders including cancer and
CC autoimmune/inflammatory, neurological, developmental, vesicle
CC trafficking, reproductive, gastrointestinal and renal disorders. These
CC disorders may include atherosclerosis, leukaemia, allergies, rheumatoid
CC arthritis, Alzheimer's disease, anxiety, diabetes, ovulatory defects,
CC renal failure and irritable bowel syndrome. A vector expressing HJNCT,
CC and an agonist of HJNCT can be used to treat or prevent a disorder
CC associated with decreased expression or activity of HJNCT. An antagonist
CC of HJNCT or a vector expressing the complement of a polynucleotide
CC encoding HJNCT can be used to treat or prevent a disorder associated
CC with increased expression or activity of HJNCT. Antibodies which bind
CC HJNCT can be used for diagnosis of disorders associated with HJNCT
CC expression or to monitor patients being treated with HJNCT, agonists,
CC antagonists or inhibitors of HJNCT. Assays are preferably carried out on
CC body fluids from a patient using radioimmunoassay, enzyme linked
CC immunosorbent assays or fluorescent activated cell sorting assays.
CC Polynucleotides encoding HJNCT are also used in hybridisation assays to
CC determine absence, presence or excess expression of HJNCT and to monitor
CC regulation of HJNCT levels during disease therapy.
XX
SQ Sequence 2742 BP; 648 A; 754 C; 682 G; 658 T; 0 other;

Query Match 100.0%; Score 690; DB 21; Length 2742;
Best Local Similarity 100.0%; Pred. No. 1e-203;
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 178

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     179 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 238

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 298

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 358

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 418

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     419 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 478

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 538

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     539 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 598

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     599 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 658

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     659 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 718

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     719 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 778

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
           |||||||||||||||||||||||||||||
Db     779 ttcaattcctacagcctgacagggtatgtg 808
```

RESULT 10
AAF93769
ID AAF93769 standard; cDNA; 2863 BP.
XX
AC AAF93769;
XX

```
DT   23-MAY-2001  (first entry)
XX
DE   Human cDNA encoding a membrane or secretory protein clone PSEC0059.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   P-PSDB; AAB88342.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 51; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 2863 BP; 670 A; 794 C; 711 G; 688 T; 0 other;

  Query Match           100.0%;  Score 690;  DB 22;  Length 2863;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-203;
  Matches  690;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     231 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 290

Qy      61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 350

Qy     121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 410

Qy     181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 470

Qy     241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     471 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 530

Qy     301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     531 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 590

Qy     361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     591 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 650

Qy     421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     651 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 710

Qy     481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     711 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 770

Qy     541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     771 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 830

Qy     601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     831 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 890

Qy     661 ttcaattcctacagcctgacagggtatgtg 690
           |||||||||||||||||||||||||||||
Db     891 ttcaattcctacagcctgacagggtatgtg 920


RESULT  11
AAS16180
ID   AAS16180 standard; cDNA; 1441 BP.
XX
AC   AAS16180;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Human CASB81 polynucleotide.
XX
KW   Human; CASB81; colorectal cancer; colon-associated disease; tumour; ss;
KW   immunogen; cytostatic; antitumour; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..769
FT                   /*tag= a
FT                   /product= "Human CASB81 protein"
FT                   /transl_except= (pos:635..637, aa:Xaa)
FT                   /note= "Xaa= unknown"
XX
PN   WO200180879-A2.
XX
PD   01-NOV-2001.
XX
PF   19-APR-2001; 2001WO-EP04456.
XX
PR   20-APR-2000; 2000GB-0009907.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Vinals De Bassols YC,  Gaulis SRJ,  Coche T,  Orntoft T;
XX
```

```
DR   WPI; 2002-041368/05.
DR   P-PSDB; AAU10521.
XX
PT   Use of a claudin 2 polypeptide and polynucleotide as vaccine for
PT   preventing and treating colorectal cancer, and as diagnostic reagents
PT   in context of the cancer  -
XX
PS   Claim 8; Page 57; 68pp; English.
XX
CC   The invention relates to the use of CASB81 polypeptides and
CC   polynucleotides for the manufacture of a medicament for treating a
CC   patient suffering from or susceptible to colorectal cancer or other
CC   colon-associated tumours or diseases. The sequences act as agents for
CC   specific prophylactic or therapeutic immunisation against tumours because
CC   they are specifically expressed or highly over-expressed in colorectal
CC   tumours compared to normal cells, and can thus be targeted by
CC   antigen-specific immune mechanisms leading to the destruction of tumour
CC   cells. They can be used in development of an antibody immunospecific for
CC   CASB81 protein, which can be used for diagnosis of tumour cells. Peptides
CC   incorporating epitopes of CASB81 provide suitable immunogens. Presence of
CC   or susceptibility to colorectal cancer can be diagnosed by analysing the
CC   presence or amount of CASB81 DNA expression or CASB81 protein activity.
CC   This sequence represents a cDNA encoding a human CASB81 polypeptide.
XX
SQ   Sequence 1441 BP; 337 A; 399 C; 357 G; 345 T; 3 other;


  Query Match             99.9%;  Score 689.6;  DB 24;  Length 1441;
  Best Local Similarity   99.9%;  Pred. No. 9.9e-204;
  Matches  689;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 136

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 196

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 256

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 316

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 376

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 436

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 496

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 556

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 616

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db    617 ctctgcttttcctgctcatsccagagaaatcgctccaactactacgatgcctaccaagcc 676

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 736

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    737 ttcaattcctacagcctgacagggtatgtg 766


RESULT  12
AAX97818
ID   AAX97818 standard; cDNA; 1524 BP.
XX
AC   AAX97818;
XX
DT   23-SEP-1999  (first entry)
XX
DE   Human secreted protein encoding cDNA #6.
XX
KW   Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW   diagnostic; gene therapy; chromosome mapping; secretion vector; ss.
XX
OS   Homo sapiens.
XX
PN   WO9925825-A2.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-IB01862.
XX
PR   04-SEP-1998;   98US-0099273.
PR   13-NOV-1997;   97US-0066677.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-347472/29.
DR   P-PSDB; AAY36134.
XX
PT   Extended cDNAs encoding secreted proteins
XX
PS   Example 28; Page 155-156; 307pp; English.
XX
CC   AAX97813-X97906 represent extended cDNA's which encode novel human
CC   secreted proteins (see AAY36129-Y36222) and which have cytostatic,
CC   thrombotic and osteopathic activity. The extended cDNAs can be used to
CC   express secreted proteins or parts of them or to obtain antibodies
CC   capable of binding to the secreted proteins. They may also be used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   Uses also include design of expression vectors and secretion vectors.
XX
SQ   Sequence 1524 BP; 365 A; 404 C; 382 G; 353 T; 20 other;


  Query Match             99.9%;  Score 689.2;  DB 20;  Length 1524;
  Best Local Similarity   99.7%;  Pred. No. 1.4e-203;
  Matches  688;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    160 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 219

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 279

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 339
```

```
Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db    340 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccakgctgcc 399

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 459

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    460 ggcatgaratgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 519

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    520 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 579

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 639

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    640 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 699

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    700 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 759

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    760 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 819

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    820 ttcaattcctacagcctgacagggtatgtg 849
```

```
RESULT 13
AAI58013
ID   AAI58013 standard; cDNA; 1540 BP.
XX
AC   AAI58013;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 216.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM38857.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 216; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1540 BP; 365 A; 419 C; 389 G; 367 T; 0 other;
```

```
  Query Match             99.8%;  Score 688.4;  DB 22;  Length 1540;
  Best Local Similarity   99.9%;  Pred. No. 2.4e-203;
  Matches  689;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 242

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 302

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 362

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 422

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 482

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 542

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    543 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 602

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 662

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    663 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 722

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
```

```
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    723 ctctgcttttcctgctcatgccagagaaatcgctccaactactacgatgcctaccaagcc 782

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    783 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 842

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    843 ttcaattcctacagcctgacagggtatgtg 872


RESULT 14
AAI59799/c
ID   AAI59799 standard; cDNA; 1761 BP.
XX
AC   AAI59799;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 3788.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM40643.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 3788; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1761 BP; 407 A; 446 C; 478 G; 430 T; 0 other;


  Query Match             99.8%;  Score 688.4;  DB 22;  Length 1761;
  Best Local Similarity   99.9%;  Pred. No. 2.6e-203;
  Matches  689;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1641 ATGGCCTCTCTTGGCCTCCAACTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGGC 1582

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1581 ACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCATT 1522

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1521 GTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGGC 1462

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1461 ATCACCCAGTGTGACATCTATAGCACCCTTCTGGGCCTGCCCGCTGACATCCAGGCTGCC 1402

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1401 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGTG 1342

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1341 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 1282

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1281 GGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCCTGTTGCCTGGAATCTT 1222

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1221 CATGGGATCCTACGGGACTTCTACTCACCACTGGTGCCTGACAGCATGAAATTTGAGATT 1162

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1161 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCATC 1102

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1101 CTCTGCTTTTCCTGCTCATGCCAGAGAAATCGCTCCAACTACTACGATGCCTACCAAGCC 1042

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1041 CAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGAG 982

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          |||||||||||||||||||||||||||||
Db    981 TTCAATTCCTACAGCCTGACAGGGTATGTG 952


RESULT 15
AAS16181
ID   AAS16181 standard; DNA; 779 BP.
XX
AC   AAS16181;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Murine CASB81 polynucleotide.
XX
KW   Mouse; CASB81; colorectal cancer; colon-associated disease; tumour; ds;
KW   immunogen; cytostatic; antitumour; gene therapy.
XX
```

OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 48..740
FT /*tag= a
FT /product= "Murine CASB81 protein"
XX
PN WO200180879-A2.
XX
PD 01-NOV-2001.
XX
PF 19-APR-2001; 2001WO-EP04456.
XX
PR 20-APR-2000; 2000GB-0009907.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Vinals De Bassols YC, Gaulis SRJ, Coche T, Orntoft T;
XX
DR WPI; 2002-041368/05.
DR P-PSDB; AAU10522.
XX
PT Use of a claudin 2 polypeptide and polynucleotide as vaccine for
PT preventing and treating colorectal cancer, and as diagnostic reagents
PT in context of the cancer -
XX
PS Claim 4; Page 57-58; 68pp; English.
XX
CC The invention relates to the use of CASB81 polypeptides and
CC polynucleotides for the manufacture of a medicament for treating a
CC patient suffering from or susceptible to colorectal cancer or other
CC colon-associated tumours or diseases. The sequences act as agents for
CC specific prophylactic or therapeutic immunisation against tumours because
CC they are specifically expressed or highly over-expressed in colorectal
CC tumours compared to normal cells, and can thus be targeted by
CC antigen-specific immune mechanisms leading to the destruction of tumour
CC cells. They can be used in development of an antibody immunospecific for
CC CASB81 protein, which can be used for diagnosis of tumour cells. Peptides
CC incorporating epitopes of CASB81 provide suitable immunogens. Presence of
CC or susceptibility to colorectal cancer can be diagnosed by analysing the
CC presence or amount of CASB81 DNA expression or CASB81 protein activity.
CC This sequence represents DNA encoding a mouse CASB81 polypeptide.
XX
SQ Sequence 779 BP; 152 A; 205 C; 216 G; 206 T; 0 other;

Query Match 79.1%; Score 546; DB 24; Length 779;
Best Local Similarity 87.0%; Pred. No. 3e-159;
Matches 600; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```
Qy      1 atggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggc 60
          |||||||| |||||| | ||||| |||||||||||||||||||| ||||||| || |||
Db     48 atggcctcccttggcgttcaactggtgggctacatcctaggccttttggggctgttaggc 107

Qy     61 acactggttgccatgctgctccccagctggaaaacaagttcttatgtcggtgccagcatt 120
          |||    |||||||||||||| |||| ||||  ||| ||||| ||||| ||||||||||||
Db    108 acatccattgccatgctgcttcccaactggcgaacgagttcctatgttggtgccagcatt 167

Qy    121 gtgacagcagttggcttctccaagggcctctggatggaatgtgccacacacagcacaggc 180
          ||||| || |||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db    168 gtgacggcggttggcttttccaagggcctctggatggagtgtgcgacacacagcacaggc 227

Qy    181 atcacccagtgtgacatctatagcacccttctgggcctgcccgctgacatccaggctgcc 240
          |||||||||| || ||||| || |||||| | || || || ||||||||||||||||||
Db    228 atcacccagtgcgatatctacagtacccttttaggacttcctgctgacatccaggctgcc 287

Qy    241 caggccatgatggtgacatccagtgcaatctcctccctggcctgcattatctctgtggtg 300
          |||||||||||||||| |||||||||| ||||| ||||| || ||||||||||||||||
Db    288 caggccatgatggtgacgtccagtgcaatgtcctcgctggcttgtattatctctgtggtg 347

Qy    301 ggcatgagatgcacagtcttctgccaggaatcccgagccaaagacagagtggcggtagca 360
          |||||||||||||| || |||||||||| || ||||| || |||||||||| ||||
Db    348 ggcatgagatgcaccgtgttctgccaggattctcgagctaaggacagagtggctgtagtg 407

Qy    361 ggtggagtctttttcatccttggaggcctcctgggattcattcctgttgcctggaatctt 420
          |||||||||||||||||||||||||| ||| ||||||| || || || ||||| |||||||||
Db    408 ggtggagtctttttcatccttggtggcatcctgggctttatcccagttgcttggaatctt 467

Qy    421 catgggatcctacgggacttctactcaccactggtgcctgacagcatgaaatttgagatt 480
          ||||| ||||| |||||||||||||| || ||||| ||||||||||||||||||||||||
Db    468 catggcatccttcgggacttctactcgccgctggttcctgacagcatgaaatttgagatt 527

Qy    481 ggagaggctctttacttgggcattatttcttccctgttctccctgatagctggaatcatc 540
          |||||||||| |||||||||| || ||  ||||||| ||  || |||| ||| |||||
Db    528 ggagaggctctgtacttgggcatcatctcagccctgttttctttggtagccggagtcatc 587

Qy    541 ctctgcttttcctgctcatcccagagaaatcgctccaactactacgatgcctaccaagcc 600
          || |||||||||||||  ||||| | |||||  |||||||||| |||| |||||| |||
Db    588 ctttgcttttcctgctcgccccagggcaatcgtaccaactactatgatggctaccaggcc 647

Qy    601 caacctcttgccacaaggagctctccaaggcctggtcaacctcccaaagtcaagagtgag 660
          || |||||||||| |||||||||||||| ||| |||||  ||||||| ||||||||||
Db    648 cagcctcttgccactaggagctctccaagatctgctcaacagcccaaagccaagagtgag 707

Qy    661 ttcaattcctacagcctgacagggtatgtg 690
          ||||| || |||||||||| |||||||||
Db    708 ttcaactcatacagcctgactgggtatgtg 737
```

Search completed: June 2, 2002, 20:09:02
Job time: 14747 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 18:58:54 ; Search time 108.28 Seconds
(without alignments)
163.332 Million cell updates/sec

Title: US-09-663-600A-139_COPY_36_107
Perfect score: 72
Sequence: 1 atggcctctcttggcctcca.....ttttgggcacactggttgcc 72

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	28	38.9	4046	1	US-07-793-961A-1	Sequence 1, Appli
c	2	28	38.9	4046	1	US-08-240-357-1	Sequence 1, Appli
	3	25.6	35.6	2104	1	US-07-984-044A-4	Sequence 4, Appli
	4	25.6	35.6	2104	1	US-08-458-393-4	Sequence 4, Appli
c	5	25.6	35.6	9551	1	US-08-056-200-93	Sequence 93, Appl
c	6	25.6	35.6	9551	2	US-08-800-644-93	Sequence 93, Appl
c	7	24.6	34.2	700	4	US-08-998-416-1131	Sequence 1131, Ap
c	8	24.4	33.9	1584	1	US-08-361-920-24	Sequence 24, Appl
c	9	24.4	33.9	1584	1	US-08-479-939-24	Sequence 24, Appl
c	10	24.4	33.9	1584	1	US-08-483-432-24	Sequence 24, Appl
	11	24.4	33.9	4919	4	US-08-456-200B-2	Sequence 2, Appli
	12	24.4	33.9	7032	2	US-08-149-097D-24	Sequence 24, Appl
	13	24.4	33.9	7032	3	US-08-949-386-24	Sequence 24, Appl
	14	24.4	33.9	7032	3	US-08-450-562-24	Sequence 24, Appl
	15	24.4	33.9	7032	4	US-08-984-709A-24	Sequence 24, Appl
	16	24.4	33.9	7089	3	US-08-949-386-25	Sequence 25, Appl
	17	24.4	33.9	7089	3	US-08-450-562-25	Sequence 25, Appl
	18	24.4	33.9	7089	4	US-08-984-709A-25	Sequence 25, Appl
	19	24.2	33.6	2226	3	US-08-889-425-1	Sequence 1, Appli
c	20	23.6	32.8	734	4	US-09-232-191-16	Sequence 16, Appl
c	21	23.6	32.8	734	4	US-09-232-200-16	Sequence 16, Appl
c	22	23.6	32.8	734	4	US-09-232-197-16	Sequence 16, Appl
c	23	23.6	32.8	734	4	US-09-232-201-16	Sequence 16, Appl
c	24	23.6	32.8	1896	4	US-09-232-200-30	Sequence 30, Appl
c	25	23.6	32.8	1896	4	US-09-232-197-30	Sequence 30, Appl
c	26	23.6	32.8	1896	4	US-09-232-201-30	Sequence 30, Appl
c	27	23.6	32.8	2907	4	US-09-232-200-52	Sequence 52, Appl
c	28	23.6	32.8	2907	4	US-09-232-197-52	Sequence 52, Appl
c	29	23.6	32.8	2907	4	US-09-232-201-52	Sequence 52, Appl
c	30	23.6	32.8	2917	4	US-09-232-200-26	Sequence 26, Appl
c	31	23.6	32.8	2917	4	US-09-232-197-26	Sequence 26, Appl
c	32	23.6	32.8	2917	4	US-09-232-201-26	Sequence 26, Appl
c	33	23.4	32.5	2343	2	US-09-018-760-3	Sequence 3, Appli
c	34	23.4	32.5	4131	3	US-08-726-214-11	Sequence 11, Appl
c	35	23.2	32.2	1516	3	US-09-059-522-4	Sequence 4, Appli
c	36	23.2	32.2	1516	3	US-09-382-027-4	Sequence 4, Appli
c	37	23.2	32.2	2057	3	US-09-059-522-2	Sequence 2, Appli
c	38	23.2	32.2	2057	3	US-09-382-027-2	Sequence 2, Appli
c	39	23	31.9	2384	1	US-07-814-964-10	Sequence 10, Appl
c	40	23	31.9	2384	1	US-08-258-442-10	Sequence 10, Appl
c	41	23	31.9	2384	1	US-08-328-809-5	Sequence 5, Appli
c	42	23	31.9	2384	5	PCT-US92-11107-10	Sequence 10, Appl
c	43	23	31.9	4821	1	US-08-053-614-3	Sequence 3, Appli
c	44	23	31.9	4821	1	US-08-316-397B-3	Sequence 3, Appli
c	45	23	31.9	4821	2	US-09-034-306-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-07-793-961A-1/c
; Sequence 1, Application US/07793961A
; Patent No. 5334521
; GENERAL INFORMATION:
; APPLICANT: Yoshihiro Ishikawa
; TITLE OF INVENTION: Cloning and Character-
; TITLE OF INVENTION: ization of a Cardiac Adenylyl Cyclase
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alan M. Gordon
; ADDRESSEE: American Cyanamid Company
; STREET: 1937 West Main Street,
; STREET: P.O. Box 60
; CITY: Stamford
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII from DW4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/793,961A
; FILING DATE: 19911118
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 31,705
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203 321 2719
; TELEFAX: 203 321 2971
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4046 base pairs listed
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-793-961A-1

Query Match 38.9%; Score 28; DB 1; Length 4046;
Best Local Similarity 66.7%; Pred. No. 0.57;

```
  Matches   40;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      13 ggcctccaacttgtgggctacatcctaggccttctggggcttttgggcacactggttgcc 72
           | ||   |||| | | |  |||||| ||  ||| |||  | ||||||| | ||||| |
Db    2008 GCCCAGGAACTCATCGACCTCATCCTCGGGGTTCAGGGCATCTTGGGCACCCCGGTTGTC 1949


RESULT   2
US-08-240-357-1/c
; Sequence 1, Application US/08240357
; Patent No. 5578481
;  GENERAL INFORMATION:
;    APPLICANT:  Ishikawa, Yoshihiro
;    TITLE OF INVENTION:  Cloning and Characterization of a
;    TITLE OF INVENTION:  Cardiac Adenylyl Cyclase
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470-8426
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/240,357
;      FILING DATE:  10-MAY-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gordon, Alan M.
;      REGISTRATION NUMBER:  30,637
;      REFERENCE/DOCKET NUMBER:  31,705-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-831-3244
;      TELEFAX:  201-831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4046 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  131..3625
US-08-240-357-1


  Query Match             38.9%;  Score 28;  DB 1;  Length 4046;
  Best Local Similarity   66.7%;  Pred. No. 0.57;
  Matches   40;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      13 ggcctccaacttgtgggctacatcctaggccttctggggcttttgggcacactggttgcc 72
           | ||   |||| | | |  |||||| ||  ||| |||  | ||||||| | ||||| |
Db    2008 GCCCAGGAACTCATCGACCTCATCCTCGGGGTTCAGGGCATCTTGGGCACCCCGGTTGTC 1949


RESULT   3
US-07-984-044A-4
; Sequence 4, Application US/07984044A
; Patent No. 5461145
;  GENERAL INFORMATION:
;    APPLICANT:  Kudo, T. et al.
;    TITLE OF INVENTION:  Sexing Method Of Bovine Embryos
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/984,044A
;      FILING DATE:  02-DEC-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7005-053
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2104 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-984-044A-4


  Query Match             35.6%;  Score 25.6;  DB 1;  Length 2104;
  Best Local Similarity   62.5%;  Pred. No. 3.3;
  Matches   40;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       4 gcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggcaca 63
           | |||||| || ||   ||||| |  ||| |  |||||   |||  |  ||    |||||
Db    1970 GGCTCTCTAGGGCTTTCACTTGGGCCCTAAAGGCTAGGTGCTCTTAGAGTTGACTGCACA 2029

Qy      64 ctgg 67
           || |
Db    2030 CTTG 2033


RESULT   4
US-08-458-393-4
; Sequence 4, Application US/08458393
; Patent No. 5661011
;  GENERAL INFORMATION:
;    APPLICANT:  Kudo, T. et al.
;    TITLE OF INVENTION:  Sexing Method Of Bovine Embryos
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,393
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/984,044
;      FILING DATE:  02-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
```

```
;      REFERENCE/DOCKET NUMBER:  7005-053
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2104 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-458-393-4


  Query Match             35.6%;  Score 25.6;  DB 1;  Length 2104;
  Best Local Similarity   62.5%;  Pred. No. 3.3;
  Matches   40;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      4 gcctctcttggcctccaacttgtgggctacatcctaggcctttctggggctttttgggcaca 63
          | |||||| || ||   ||||| |  ||| |  |||||   |||  |  ||    |||||
Db   1970 GGCTCTCTAGGGCTTTCACTTGGGCCCTAAAGGCTAGGTGCTCTTAGAGTTGACTGCACA 2029

Qy     64 ctgg 67
          || |
Db   2030 CTTG 2033


RESULT   5
US-08-056-200-93/c
; Sequence 93, Application US/08056200
; Patent No. 5616500
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/056,200
;      FILING DATE:  30-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fedrick, Michael F.
;      REGISTRATION NUMBER:  36,799
;      REFERENCE/DOCKET NUMBER:  NIH054.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  93:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9551 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1507..1644
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1645..2511
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2512..8070
US-08-056-200-93


  Query Match             35.6%;  Score 25.6;  DB 1;  Length 9551;
  Best Local Similarity   62.5%;  Pred. No. 4.8;
  Matches   40;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      2 tggcctctcttggcctccaacttgtgggctacatcctaggcctttctggggctttttgggca 61
          || |  ||| |  |||    ||  |||    | ||||   ||| |||| ||| |||||||
Db   5021 TGTCGCCTCCTCTCCTGATCCTCCTGGAGGCCGTCCTCCTCCTCCTGGAGCTGTTGGGCA 4962

Qy     62 cact 65
          | ||
Db   4961 CGCT 4958


RESULT   6
US-08-800-644-93/c
; Sequence 93, Application US/08800644
; Patent No. 5958752
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/800,644
;      FILING DATE:  14-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/056,200
;      FILING DATE:  30-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fedrick, Michael F.
;      REGISTRATION NUMBER:  36,799
;      REFERENCE/DOCKET NUMBER:  NIH054.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  93:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9551 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
```

```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1507..1644
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1645..2511
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2512..8070
US-08-800-644-93


  Query Match             35.6%;  Score 25.6;  DB 2;  Length 9551;
  Best Local Similarity   62.5%;  Pred. No. 4.8;
  Matches   40;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 tggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggca 61
           || |  ||| |  |||   ||  |||    | ||||   ||| |||| ||| |||||||
Db    5021 TGTCGCCTCCTCTCCTGATCCTCCTGGAGGCCGTCCTCCTCCTCCTGGAGCTGTTGGGCA 4962

Qy      62 cact 65
           | ||
Db    4961 CGCT 4958


RESULT   7
US-08-998-416-1131/c
; Sequence 1131, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  1131:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1689RP
US-08-998-416-1131


  Query Match             34.2%;  Score 24.6;  DB 4;  Length 700;
  Best Local Similarity   61.9%;  Pred. No. 5.7;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 tggcctctcttggcctccaacttgtgggctacatcctaggccttctggggcttttgggca 61
           ||| |   ||||  |   ||||||| ||  || ||| |  ||| ||  ||| |   ||||
Db     609 TGGACGGGCTTGTTCATGAACTTGTCGGTCACGTCCAACACCTCCTTTGGCATGGTGGCA 550

Qy      62 cac 64
           ||
Db     549 GAC 547


RESULT   8
US-08-361-920-24/c
; Sequence 24, Application US/08361920
; Patent No. 5457046
;  GENERAL INFORMATION:
;    APPLICANT:  Woeldike, Helle F.
;    APPLICANT:  Hagen, Frederick
;    APPLICANT:  Hjort, Carsten M.
;    APPLICANT:  Sven, Hastrup
;    TITLE OF INVENTION:  An Enzyme Capable of Degrading Cellulose
;    TITLE OF INVENTION:  or Hemicellulose
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5457046o No. 5457046disk of No. 5457046th America, Inc.
;      STREET:  405 Lexington Avenue, 62nd Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/361,920
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/940,860
;      FILING DATE:  28-OCT-1992
;      APPLICATION NUMBER:  DK 1158/90
;      FILING DATE:  09-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/DK91/00124
;      FILING DATE:  08-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3435.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-867-0298
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1584 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Fusarium oxysporum
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  55..1335
US-08-361-920-24


  Query Match             33.9%;  Score 24.4;  DB 1;  Length 1584;
  Best Local Similarity   63.8%;  Pred. No. 8.1;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      4 gcctctcttggcctccaacttgtgggctacatcctaggccttctggggctttgggca 61
          ||  |  | || ||| | |||| ||      |||| |||||| | |||  |||| |||
Db    162 GCAGCGGTAGGTCTCGAGCTTGGGGTGCTGCTCCTTGGCCTTGTCGGGGGTTTGAGCA 105


RESULT   9
US-08-479-939-24/c
; Sequence 24, Application US/08479939
; Patent No. 5686593
;  GENERAL INFORMATION:
;    APPLICANT:  Woeldike, Helle F.
;    APPLICANT:  Hagen, Frederick
;    APPLICANT:  Hjort, Carsten M.
;    APPLICANT:  Sven, Hastrup
;    TITLE OF INVENTION:  An Enzyme Capable of Degrading Cellulose
;    TITLE OF INVENTION:  or Hemicellulose
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5686593o No. 5686593disk of No. 5686593th America, Inc.
;      STREET:  405 Lexington Avenue, 62nd Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,939
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/361,920
;      FILING DATE:  22-DEC-1994
;      APPLICATION NUMBER:  US 07/940,860
;      FILING DATE:  28-OCT-1992
;      APPLICATION NUMBER:  DK 1158/90
;      FILING DATE:  09-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/DK91/00124
;      FILING DATE:  08-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3435.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-867-0298
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1584 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Fusarium oxysporum
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..1335
US-08-479-939-24
```

```
  Query Match             33.9%;  Score 24.4;  DB 1;  Length 1584;
  Best Local Similarity   63.8%;  Pred. No. 8.1;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      4 gcctctcttggcctccaacttgtgggctacatcctaggccttctggggctttgggca 61
          ||  |  | || ||| | |||| ||      |||| |||||| | |||  |||| |||
Db    162 GCAGCGGTAGGTCTCGAGCTTGGGGTGCTGCTCCTTGGCCTTGTCGGGGGTTTGAGCA 105


RESULT  10
US-08-483-432-24/c
; Sequence 24, Application US/08483432
; Patent No. 5763254
;  GENERAL INFORMATION:
;    APPLICANT:  Woeldike, Helle F.
;    APPLICANT:  Hagen, Frederick
;    APPLICANT:  Hjort, Carsten M.
;    APPLICANT:  Sven, Hastrup
;    TITLE OF INVENTION:  An Enzyme Capable of Degrading Cellulose
;    TITLE OF INVENTION:  or Hemicellulose
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5763254o No. 5763254disk of No. 5763254th America, Inc.
;      STREET:  405 Lexington Avenue, 62nd Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,432
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/361,920
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/940,860
;      FILING DATE:  28-OCT-1992
;      APPLICATION NUMBER:  DK 1158/90
;      FILING DATE:  09-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/DK91/00124
;      FILING DATE:  08-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3435.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-867-0298
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1584 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Fusarium oxysporum
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..1335
US-08-483-432-24


  Query Match             33.9%;  Score 24.4;  DB 1;  Length 1584;
  Best Local Similarity   63.8%;  Pred. No. 8.1;
```

```
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       4 gcctctcttggcctccaacttgtgggctacatcctaggccttctggggctttgggca 61
           ||  |  | || ||| | |||| ||      |||| |||||| | |||  |||| |||
Db     162 GCAGCGGTAGGTCTCGAGCTTGGGGTGCTGCTCCTTGGCCTTGTCGGGGGTTTGAGCA 105


RESULT  11
US-08-456-200B-2
; Sequence 2, Application US/08456200B
; Patent No. 6229000
;  GENERAL INFORMATION:
;    APPLICANT:  Franz, Jurgen; Weingartner, Bernhard;
;    APPLICANT:  Unterbeck, Axel; Rae, Peter
;    TITLE OF INVENTION:  TISSUE-SPECIFIC HUMAN NEURONAL
;    TITLE OF INVENTION:  CALCIUM CHANNEL SUB-TYPES AND
;    TITLE OF INVENTION:  THEIR USE
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  NEC Powermate SX/20
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,200B
;      FILING DATE:  31-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/094,712
;      FILING DATE:  19-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/858,278
;      FILING DATE:  26-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/064,778
;      FILING DATE:  19-MAY-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 41 10 785
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 8398.3-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4919 nucleotides
;      TYPE:  Nucleotide
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  cDNA
US-08-456-200B-2


  Query Match             33.9%;  Score 24.4;  DB 4;  Length 4919;
  Best Local Similarity   63.8%;  Pred. No. 11;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      15 cctccaacttgtgggctacatcctaggccttctggggcttttgggcacactggttgcc 72
           | ||| |    || |  |||||||  | | || ||| | |  | ||| || |||| |||
Db    2566 CTTCCGAGACTTGTGGAACATCCTGGACTTTGTGGTGGTCGTTGGCGCATTGGTGGCC 2623


RESULT  12
US-08-149-097D-24
; Sequence 24, Application US/08149097D
; Patent No. 5874236
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  Feldman, Daniel
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Brenner, Robert
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,097D
;      FILING DATE:  05-NOV-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/105,536
;      FILING DATE:  11-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US92/06903
;      FILING DATE:  14-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/914,231
;      FILING DATE:  13-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/868,354
;      FILING DATE:  10-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/745,206
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/620,250
;      FILING DATE:  30-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/482,384
;      FILING DATE:  20-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/603,751
;      FILING DATE:  04-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US89/01408
;      FILING DATE:  04-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/176,899
;      FILING DATE:  04-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  6362-55038
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 238-0999
;      TELEFAX:  (619) 238-0062
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7032 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  169..6921
;       OTHER INFORMATION:  /product= "AlphaIE-1 subunit of
;       OTHER INFORMATION:  human calcium channel"
US-08-149-097D-24


  Query Match             33.9%;  Score 24.4;  DB 2;  Length 7032;
  Best Local Similarity   63.8%;  Pred. No. 12;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      15 cctccaacttgtgggctacatcctaggccttctggggcttttgggcacactggttgcc 72
           | ||| |    || |  ||||||| | | || ||| | |  | ||| || |||| |||
Db    3759 CTTCCGAGACTTGTGGAACATCCTGGACTTTGTGGTGGTCGTTGGCGCATTGGTGGCC 3816


RESULT  13
US-08-949-386-24
; Sequence 24, Application US/08949386
; Patent No. 6090623
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Gillespie, Alison
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/949,386
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,012
;      FILING DATE:  11-AUG-1994
;      APPLICATION NUMBER:  08/149,097
;      FILING DATE:  5-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/105,536
;      FILING DATE:  11-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  519808
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 238-0999
;      TELEFAX:  (619) 238-0062
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7032 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  166..6921
;      OTHER INFORMATION:  /standard_name= "Alpha-1E-1"
US-08-949-386-24


  Query Match             33.9%;  Score 24.4;  DB 3;  Length 7032;
  Best Local Similarity   63.8%;  Pred. No. 12;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      15 cctccaacttgtgggctacatcctaggccttctggggcttttgggcacactggttgcc 72
           | ||| |    || |  ||||||| | | || ||| | |  | ||| || |||| |||
Db    3759 CTTCCGAGACTTGTGGAACATCCTGGACTTTGTGGTGGTCGTTGGCGCATTGGTGGCC 3816


RESULT  14
US-08-450-562-24
; Sequence 24, Application US/08450562
; Patent No. 6096514
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Gillespie, Alison
;    APPLICANT:  Feldman, Daniel
;    APPLICANT:  Brenner, Robert
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/450,562
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/404,950
;      FILING DATE:  13-MAR-1995
;      APPLICATION NUMBER:  08/336,257
;      FILING DATE:  7-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/314,083
;      FILING DATE:  28-SEPT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/311,363
;      FILING DATE:  23-SEPT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/290,012
;      FILING DATE:  11-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/223,305
;      FILING DATE:  4-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/193,078
;      FILING DATE:  07-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,097
;      FILING DATE:  5-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/105,536
;      FILING DATE:  11-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/914,231
;      FILING DATE:  13-JULY-1992
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/868,354
;      FILING DATE:  10-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06903
;      FILING DATE:  14-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/745,206
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/620,250
;      FILING DATE:  30-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/603,751
;      FILING DATE:  08-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/482,384
;      FILING DATE:  02-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US89/01408
;      FILING DATE:  04-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/176,899
;      FILING DATE:  04-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  6362-519812
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 238-0999
;      TELEFAX:  (619) 238-0062
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7032 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  166..6921
;      OTHER INFORMATION:  /standard_name= "Alpha-1E-1"
US-08-450-562-24


  Query Match             33.9%;  Score 24.4;  DB 3;  Length 7032;
  Best Local Similarity   63.8%;  Pred. No. 12;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      15 cctccaacttgtgggctacatcctaggccttctggggcttttgggcacactggttgcc 72
           | ||| |    || |  ||||||| | | || ||| | |   | ||| || |||| |||
Db    3759 CTTCCGAGACTTGTGGAACATCCTGGACTTTGTGGTGGTCGTTGGCGCATTGGTGGCC 3816


RESULT  15
US-08-984-709A-24
; Sequence 24, Application US/08984709A
; Patent No. 6320032
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Mark E.
;    APPLICANT:  Stauderman, Kenneth A.
;    APPLICANT:  Harpold, Michael M.
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Heller Ehrman White & McAuliffe
;      STREET:  4250 Executive Square, Suite 700
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/984,709A
;      FILING DATE:  02-DEC-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  24735-9815 (formerly 6362-9815)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 450-8400
;      TELEFAX:  (619) 587-5360
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7032 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  166..6921
;      OTHER INFORMATION:  /standard_name= "Alpha-1E-1"
US-08-984-709A-24


  Query Match             33.9%;  Score 24.4;  DB 4;  Length 7032;
  Best Local Similarity   63.8%;  Pred. No. 12;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      15 cctccaacttgtgggctacatcctaggccttctggggcttttgggcacactggttgcc 72
           | ||| |    || |  ||||||| | | || ||| | |   | ||| || |||| |||
Db    3759 CTTCCGAGACTTGTGGAACATCCTGGACTTTGTGGTGGTCGTTGGCGCATTGGTGGCC 3816


Search completed:  June  2, 2002, 18:58:58
Job time:  11083 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:59:00 ; Search time 64.27 Seconds
(without alignments)
87.411 Million cell updates/sec

Title: US-09-663-600A-186
Perfect score: 1174
Sequence: 1 MASLGLQLVGYILGLLGLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	223.5	19.0	137	4	US-09-188-930-174	Sequence 174, App
2	185.5	15.8	207	1	US-08-557-917A-2	Sequence 2, Appli
3	185.5	15.8	207	4	US-09-084-153-2	Sequence 2, Appli
4	185.5	15.8	207	4	US-09-084-079-2	Sequence 2, Appli
5	179	15.2	218	4	US-09-084-079-5	Sequence 5, Appli
6	91	7.8	160	3	US-08-600-430-5	Sequence 5, Appli
7	87	7.4	160	3	US-08-600-430-6	Sequence 6, Appli
8	87	7.4	161	1	US-08-557-917A-3	Sequence 3, Appli
9	87	7.4	161	4	US-09-084-153-3	Sequence 3, Appli
10	87	7.4	161	4	US-09-084-079-3	Sequence 3, Appli
11	85	7.2	681	6	5194595-19	Patent No. 5194595
12	84.5	7.2	580	2	US-08-677-049-12	Sequence 12, Appl
13	84	7.2	521	2	US-08-406-855A-19	Sequence 19, Appl
14	84	7.2	521	3	US-09-206-899-19	Sequence 19, Appl
15	83	7.1	144	1	US-08-518-474-8	Sequence 8, Appli
16	82	7.0	157	2	US-08-883-070-3	Sequence 3, Appli
17	82	7.0	468	2	US-08-390-000A-7	Sequence 7, Appli
18	82	7.0	477	1	US-08-444-734A-4	Sequence 4, Appli
19	82	7.0	477	1	US-08-087-772A-16	Sequence 16, Appl
20	80.5	6.9	314	3	US-08-988-876-7	Sequence 7, Appli
21	80.5	6.9	417	1	US-08-553-888A-3	Sequence 3, Appli
22	79	6.7	158	3	US-09-010-809-19	Sequence 19, Appl
23	78.5	6.7	277	1	US-08-118-270-68	Sequence 68, Appl
24	78.5	6.7	277	5	PCT-US93-08528-68	Sequence 68, Appl
25	78.5	6.7	381	4	US-08-876-798A-2	Sequence 2, Appli
26	78	6.6	438	2	US-08-677-049-9	Sequence 9, Appli
27	77.5	6.6	381	1	US-08-192-288-2	Sequence 2, Appli
28	77.5	6.6	381	2	US-08-687-355A-2	Sequence 2, Appli
29	77.5	6.6	381	4	US-09-200-673-16	Sequence 16, Appl
30	77	6.6	160	2	US-08-883-070-7	Sequence 7, Appli
31	76.5	6.5	475	1	US-07-686-591-4	Sequence 4, Appli
32	76.5	6.5	475	1	US-07-970-715-4	Sequence 4, Appli
33	75.5	6.4	227	1	US-08-254-493-1	Sequence 1, Appli
34	75.5	6.4	227	2	US-08-253-751-6	Sequence 6, Appli
35	75.5	6.4	227	2	US-08-453-925-6	Sequence 6, Appli
36	75.5	6.4	227	4	US-08-403-253A-6	Sequence 6, Appli
37	75.5	6.4	228	1	US-08-408-222B-1	Sequence 1, Appli
38	75.5	6.4	360	4	US-09-045-583-51	Sequence 51, Appl
39	75.5	6.4	501	1	US-08-722-001-14	Sequence 14, Appl
40	75.5	6.4	501	2	US-08-467-568-9	Sequence 9, Appli
41	75.5	6.4	501	2	US-09-030-582-9	Sequence 9, Appli
42	75.5	6.4	572	5	PCT-US91-08177-11	Sequence 11, Appl
43	75.5	6.4	647	1	US-08-218-943-1	Sequence 1, Appli
44	74.5	6.3	330	3	US-08-115-753-1	Sequence 1, Appli
45	74.5	6.3	419	3	US-08-115-753-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-188-930-174
; Sequence 174, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 174
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Human
US-09-188-930-174

Query Match 19.0%; Score 223.5; DB 4; Length 137;
Best Local Similarity 35.7%; Pred. No. 4.2e-17;
Matches 41; Conservative 30; Mismatches 43; Indels 1; Gaps 1;

```
Qy      67 YSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKD-RVAVAGGVFF 125
           | ::| | | :|| :|:|| |  :  ||  :: :||:||    : : |  |:|: ||: |
Db       7 YDSVLALSAALQATRALMVVSLVLGFLAMFVATMGMKCTRCGGDDKVKKARIAMGGGIIF 66

Qy     126 ILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           |: ||   :  :|  | |: |||:||:| ::|:| | |:::|   |   :: | :
Db      67 IVAGLAALVACSWYGHQIVTDFYNPLIPTNIKYEFGPAIFIGWAGSALVILGGAL 121
```

RESULT 2
US-08-557-917A-2
; Sequence 2, Application US/08557917A
; Patent No. 5756300
; GENERAL INFORMATION:
; APPLICANT: Bronstein, Jeff M.
; APPLICANT: Seitz, Robert S.
; APPLICANT: Lallone, Roger L.
; TITLE OF INVENTION: Oligodendrocyte-specific Protein and Method for
; TITLE OF INVENTION: Diagnosing and Treating Disease
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak

```
;       STREET:  225 S. Lake Avenue, 9th Floor
;       CITY:  Pasadena
;       STATE:  California
;       ZIP:  91101
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  Windows version 3.11
;       SOFTWARE:  WordPerfect for Windows version 6.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/557,917A
;       FILING DATE:  14-NOVEMBER-1995
;       CLASSIFICATION:  436
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Farah, David A.
;       REGISTRATION NUMBER:  38,134
;       REFERENCE/DOCKET NUMBER:  11201
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (818)796-4000
;       TELEFAX:  (818)795-6321
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  207 amino acid residues
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-557-917A-2


  Query Match             15.8%;  Score 185.5;  DB 1;  Length 207;
  Best Local Similarity   28.7%;  Pred. No. 1.2e-12;
  Matches   64;  Conservative   35;  Mismatches   91;  Indels   33;  Gaps   10;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSW-KTSSYVGASIVTA-----VGFSKGLWMEC 54
           | :  ||:||::   :| :| :|       |  | ||   :| |      :| ||||| :|
Db       1 MVATCLQVVGFVTSFVGWIGIIVTTSTNDWVVTCSY---TIPTCRKMDELG-SKGLWADC 56

Qy      55 ATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQE-SRA 113
             :||:  |     :| ||  :|| :|:|: :| :   | :: :  : |     |   |
Db      57 VM-ATGLYHCKPLVDILILPGYVQACRALMIAASVLGLPAILLLLTVLPCIRMGHEPGVA 115

Qy     114 KDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI-----GEALYLGI 168
           | | |   ||  ||  |   :   |              |     ||     | :|| |
Db     116 KYRRAQLAGVLLILLALCAIVATIW------------FPVCAHREITIVSFGYSLYAGW 162

Qy     169 ISSLFSLIAG-IILCFSCSSQRNRSN--YYDAYQAQPLATRSS 208
           | ::  |: | :|:| |  :|    |  || :  : |   :|:
Db     163 IGAVMCLVGGCVIVCCSGDAQSFGENRFYYSSGSSSPTHAKSA 205


RESULT   3
US-09-084-153-2
; Sequence 2, Application US/09084153
; Patent No. 6147191
;  GENERAL INFORMATION:
;    APPLICANT:  Bronstein, Jeff M.
;    TITLE OF INVENTION:  Oligodendrocyte-specific Protein
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 S. Lake Avenue, 9th Floor
;      CITY:  Pasadena
;      STATE:  California
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows version 8.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,153
;      FILING DATE:  26-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farah, David A.
;      REGISTRATION NUMBER:  38,134
;      REFERENCE/DOCKET NUMBER:  11201-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (626)796-4000
;      TELEFAX:  (626)795-6321
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  207 amino acid residues
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-153-2


  Query Match             15.8%;  Score 185.5;  DB 4;  Length 207;
  Best Local Similarity   28.7%;  Pred. No. 1.2e-12;
  Matches   64;  Conservative   35;  Mismatches   91;  Indels   33;  Gaps   10;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSW-KTSSYVGASIVTA-----VGFSKGLWMEC 54
           | :  ||:||::   :| :| :|       |  | ||   :| |      :| ||||| :|
Db       1 MVATCLQVVGFVTSFVGWIGIIVTTSTNDWVVTCSY---TIPTCRKMDELG-SKGLWADC 56

Qy      55 ATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQE-SRA 113
             :||:  |     :| ||  :|| :|:|: :| :   | :: :  : |     |   |
Db      57 VM-ATGLYHCKPLVDILILPGYVQACRALMIAASVLGLPAILLLLTVLPCIRMGHEPGVA 115

Qy     114 KDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI-----GEALYLGI 168
           | | |   ||  ||  |   :   |              |     ||     | :|| |
Db     116 KYRRAQLAGVLLILLALCAIVATIW------------FPVCAHREITIVSFGYSLYAGW 162

Qy     169 ISSLFSLIAG-IILCFSCSSQRNRSN--YYDAYQAQPLATRSS 208
           | ::  |: | :|:| |  :|    |  || :  : |   :|:
Db     163 IGAVMCLVGGCVIVCCSGDAQSFGENRFYYSSGSSSPTHAKSA 205


RESULT   4
US-09-084-079-2
; Sequence 2, Application US/09084079
; Patent No. 6150136
;  GENERAL INFORMATION:
;    APPLICANT:  Bronstein, Jeff M.
;    APPLICANT:  Seitz, Robert S.
;    APPLICANT:  Lallone, Roger L.
;    TITLE OF INVENTION:  Oligodendrocyte-specific Protein and Method for
;    TITLE OF INVENTION:  Diagnosing and Treating Disease
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 S. Lake Avenue, 9th Floor
;      CITY:  Pasadena
;      STATE:  California
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows version 8.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,079
;      FILING DATE:  22-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farah, David A.
;      REGISTRATION NUMBER:  38,134
;      REFERENCE/DOCKET NUMBER:  11201-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (626)796-4000
;      TELEFAX:  (626)795-6321
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  207 amino acid residues
```

```
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-079-2


  Query Match             15.8%;  Score 185.5;  DB 4;  Length 207;
  Best Local Similarity   28.7%;  Pred. No. 1.2e-12;
  Matches   64;  Conservative   35;  Mismatches   91;  Indels   33;  Gaps   10;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSW-KTSSYVGASIVTA-----VGFSKGLWMEC 54
           | :  ||:||::   :| :| :|      |  | ||   :| |      :| ||||| :|
Db       1 MVATCLQVVGFVTSFVGWIGIIVTTSTNDWVVTCSY---TIPTCRKMDELG-SKGLWADC 56

Qy      55 ATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQE-SRA 113
             :||:  |     :| ||  :|| :|:|: :| :   | :: :  : |      |   |
Db      57 VM-ATGLYHCKPLVDILILPGYVQACRALMIAASVLGLPAILLLLTVLPCIRMGHEPGVA 115

Qy     114 KDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEI-----GEALYLGI 168
           | | |   ||  ||  |   :   |               |     ||     | :|| |
Db     116 KYRRAQLAGVLLILLALCAIVATIW------------FPVCAHREITIVSFGYSLYAGW 162

Qy     169 ISSLFSLIAG-IILCFSCSSQRNRSN--YYDAYQAQPLATRSS 208
           | ::  |: | :|:| |  :|    |  || :  : |   :|:
Db     163 IGAVMCLVGGCVIVCCSGDAQSFGENRFYYSSGSSSPTHAKSA 205


RESULT   5
US-09-084-079-5
; Sequence 5, Application US/09084079
; Patent No. 6150136
;  GENERAL INFORMATION:
;    APPLICANT:  Bronstein, Jeff M.
;    APPLICANT:  Seitz, Robert S.
;    APPLICANT:  Lallone, Roger L.
;    TITLE OF INVENTION:  Oligodendrocyte-specific Protein and Method for
;    TITLE OF INVENTION:  Diagnosing and Treating Disease
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 S. Lake Avenue, 9th Floor
;      CITY:  Pasadena
;      STATE:  California
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows version 8.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,079
;      FILING DATE:  22-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farah, David A.
;      REGISTRATION NUMBER:  38,134
;      REFERENCE/DOCKET NUMBER:  11201-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (626)796-4000
;      TELEFAX:  (626)795-6321
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  218 amino acid residues
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-079-5


  Query Match             15.2%;  Score 179;  DB 4;  Length 218;
  Best Local Similarity   27.2%;  Pred. No. 6.5e-12;
  Matches   62;  Conservative   39;  Mismatches  107;  Indels   20;  Gaps    8;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTA-----VGFSKGLWMECA 55
           | :  ||:||::   :| :| :|      |  :   | :| |      :| ||||| :|
Db       1 MVATCLQVVGFVTSFVGWIGVIVTTSTNDWVVT--CGYTIPTCRKLDELG-SKGLWADCV 57

Qy      56 THSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQE-SRAK 114
             :||:  |     :| ||  :|| :|:|: :| :   | :: :  : |    ||   ||
Db      58 M-ATGLYHCKPLVDILILPGYVQACRALMIAASVLGLPAILLLLTVLPCIRMGQEPGVAK 116

Qy     115 DRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFS 174
           | |   ||  ||  |   :   |         |      ::     | :|| | | ::
Db     117 YRRAQLAGVLLILLALCALVATIW--------FPVCAHRETTIVSFGYSLYAGWIGAVLC 168

Qy     175 LIAG-IILCFSCSSQRNRSNYYDAYQA-QPLATRSSPRPGQPPKVKSE 220
           |: | :||| :  :|    |     :|  |   | |   :  :: :|
Db     169 LVGGCVILCCAGDAQAFGENVSTTLRALAPRLMRRVPTYKRAARLPTE 216


RESULT   6
US-08-600-430-5
; Sequence 5, Application US/08600430
; Patent No. 6124436
;  GENERAL INFORMATION:
;    APPLICANT:  McClanahan, Terrill K.
;    APPLICANT:  Gorman, Daniel M.
;    APPLICANT:  Bolin, Laurel M.
;    TITLE OF INVENTION:  PURIFIED MAMMALIAN MONOCYTE ANTIGENS AND
;    TITLE OF INVENTION:  RELATED REAGENTS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/600,430
;      FILING DATE:  13-FEB-1996
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0517
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-852-9196
;      TELEFAX:  415-496-1200
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  160 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-600-430-5


  Query Match             7.8%;  Score 91;  DB 3;  Length 160;
  Best Local Similarity   21.7%;  Pred. No. 0.021;
  Matches   41;  Conservative   33;  Mismatches   69;  Indels   46;  Gaps    8;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           |  | | ::   : :| ||   |: ::  |          :   | :  ||  |:| |:|
Db       1 MLLLLLSIIVLHVAVLVLL--FVSTIVSQW----------IVGNGHATDLWQNCSTSSSG 48

Qy      61 ITQCDIYSTLLGLPAD-IQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAV 119
                :::      | : :|: || |: |   |    |:|:    | :|      :  :
Db      49 ----NVHHCFSSSPNEWLQSVQATMILSIIFS----ILSLFLFFCQLFTLTKGGRFYIT- 99
```

```
Qy    120 AGGVFFILGGLLGFIPVA--------WNLHGILRDFYSPLVPDSMKFEIGEALYLGIISS 171
          :|:| || ||      |        |:|:                :  | |  |  ::
Db    100 --GIFQILAGLCVMSAAAIYTVRHPEWHLNS-------------DYSYGFAYILAWVAF 143

Qy    172 LFSLIAGII 180
            :|::|:|
Db    144 PLALLSGVI 152


RESULT   7
US-08-600-430-6
; Sequence 6, Application US/08600430
; Patent No. 6124436
;  GENERAL INFORMATION:
;    APPLICANT:  McClanahan, Terrill K.
;    APPLICANT:  Gorman, Daniel M.
;    APPLICANT:  Bolin, Laurel M.
;    TITLE OF INVENTION:  PURIFIED MAMMALIAN MONOCYTE ANTIGENS AND
;    TITLE OF INVENTION:  RELATED REAGENTS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/600,430
;      FILING DATE:  13-FEB-1996
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0517
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-852-9196
;      TELEFAX:  415-496-1200
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  160 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-600-430-6


  Query Match             7.4%;  Score 87;  DB 3;  Length 160;
  Best Local Similarity   25.8%;  Pred. No. 0.058;
  Matches   46;  Conservative   26;  Mismatches   70;  Indels   36;  Gaps    8;

Qy     12 ILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTLL 71
          :| |||:|   :|:|:  :  |: |   :|   | :  ||  | | : |  |   ||:
Db      2 LLLLLGILFLHIAVLVLLF-VSTIVSQWLV-GNGHTTDLWQNCTTSALGAVQ-HCYSS-- 56

Qy     72 GLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQ-ESRAKDRVAVAGGVFFILGGL 130
          :   :|: || |: |   | ||  :         |||  :  |       | | || ||
Db     57 SVSEWLQSVQATMILSVIFSVLALFL--------FFCQLFTLTKGGRFYITGFFQILAGL 108

Qy    131 LGFIPVA--------WNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
                |        |::             :  :  | |  |  ::   :|::|||
Db    109 CVMSAAAIYTVRHSEWHV-------------NTDYSYGFAYILAWVAFPLALLSGII 152


RESULT   8
US-08-557-917A-3
; Sequence 3, Application US/08557917A
; Patent No. 5756300
;  GENERAL INFORMATION:
;    APPLICANT:  Bronstein, Jeff M.
;    APPLICANT:  Seitz, Robert S.
;    APPLICANT:  Lallone, Roger L.
;    TITLE OF INVENTION:  Oligodendrocyte-specific Protein and Method for
;    TITLE OF INVENTION:  Diagnosing and Treating Disease
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 S. Lake Avenue, 9th Floor
;      CITY:  Pasadena
;      STATE:  California
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Windows version 3.11
;      SOFTWARE:  WordPerfect for Windows version 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/557,917A
;      FILING DATE:  14-NOVEMBER-1995
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farah, David A.
;      REGISTRATION NUMBER:  38,134
;      REFERENCE/DOCKET NUMBER:  11201
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (818)796-4000
;      TELEFAX:  (818)795-6321
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  161 amino acid residues
;      TYPE:  amino acids
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-557-917A-3


  Query Match             7.4%;  Score 87;  DB 1;  Length 161;
  Best Local Similarity   25.8%;  Pred. No. 0.058;
  Matches   46;  Conservative   26;  Mismatches   70;  Indels   36;  Gaps    8;

Qy     12 ILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTLL 71
          :| |||:|   :|:|:  :  |: |   :|   | :  ||  | | : |  |   ||:
Db      2 LLLLLGILFLHIAVLVLLF-VSTIVSQWLV-GNGHTTDLWQNCTTSALGAVQ-HCYSS-- 56

Qy     72 GLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQ-ESRAKDRVAVAGGVFFILGGL 130
          :   :|: || |: |   | ||  :         |||  :  |       | | || ||
Db     57 SVSEWLQSVQATMILSVIFSVLALFL--------FFCQLFTLTKGGRFYITGFFQILAGL 108

Qy    131 LGFIPVA--------WNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
                |        |::             :  :  | |  |  ::   :|::|||
Db    109 CVMSAAAIYTVRHSEWHV-------------NTDYSYGFAYILAWVAFPLALLSGII 152


RESULT   9
US-09-084-153-3
; Sequence 3, Application US/09084153
; Patent No. 6147191
;  GENERAL INFORMATION:
;    APPLICANT:  Bronstein, Jeff M.
;    TITLE OF INVENTION:  Oligodendrocyte-specific Protein
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 S. Lake Avenue, 9th Floor
;      CITY:  Pasadena
;      STATE:  California
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM compatible
```

```
;       OPERATING SYSTEM:  Windows 95
;       SOFTWARE:  WordPerfect for Windows version 8.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/084,153
;       FILING DATE:  26-MAY-1998
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Farah, David A.
;       REGISTRATION NUMBER:  38,134
;       REFERENCE/DOCKET NUMBER:  11201-2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (626)796-4000
;       TELEFAX:  (626)795-6321
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  161 amino acid residues
;       TYPE:  amino acids
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-084-153-3


  Query Match             7.4%;  Score 87;  DB 4;  Length 161;
  Best Local Similarity   25.8%;  Pred. No. 0.058;
  Matches   46;  Conservative   26;  Mismatches   70;  Indels   36;  Gaps    8;

Qy      12 ILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTLL 71
           :| |||:|   :|:|:  :  |: |   :|   | :  ||  | | : |  |   ||:
Db       2 LLLLLGILFLHIAVLVLLF-VSTIVSQWLV-GNGHTTDLWQNCTTSALGAVQ-HCYSS-- 56

Qy      72 GLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQ-ESRAKDRVAVAGGVFFILGGL 130
             :   :|: || |: |   | ||  :         |||  :  |       | | || ||
Db      57 SVSEWLQSVQATMILSVIFSVLALFL--------FFCQLFTLTKGGRFYITGFFQILAGL 108

Qy     131 LGFIPVA--------WNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
                |        |::              :  :  | |  |  ::   :|::|||
Db     109 CVMSAAAIYTVRHSEWHV--------------NTDYSYGFAYILAWVAFPLALLSGII 152


RESULT  10
US-09-084-079-3
; Sequence 3, Application US/09084079
; Patent No. 6150136
;  GENERAL INFORMATION:
;    APPLICANT:  Bronstein, Jeff M.
;    APPLICANT:  Seitz, Robert S.
;    APPLICANT:  Lallone, Roger L.
;    TITLE OF INVENTION:  Oligodendrocyte-specific Protein and Method for
;    TITLE OF INVENTION:  Diagnosing and Treating Disease
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 S. Lake Avenue, 9th Floor
;      CITY:  Pasadena
;      STATE:  California
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows version 8.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,079
;      FILING DATE:  22-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farah, David A.
;      REGISTRATION NUMBER:  38,134
;      REFERENCE/DOCKET NUMBER:  11201-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (626)796-4000
;      TELEFAX:  (626)795-6321
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  161 amino acid residues
;      TYPE:  amino acids
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-079-3


  Query Match             7.4%;  Score 87;  DB 4;  Length 161;
  Best Local Similarity   25.8%;  Pred. No. 0.058;
  Matches   46;  Conservative   26;  Mismatches   70;  Indels   36;  Gaps    8;

Qy      12 ILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTLL 71
           :| |||:|   :|:|:  :  |: |   :|   | :  ||  | | : |  |   ||:
Db       2 LLLLLGILFLHIAVLVLLF-VSTIVSQWLV-GNGHTTDLWQNCTTSALGAVQ-HCYSS-- 56

Qy      72 GLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQ-ESRAKDRVAVAGGVFFILGGL 130
             :   :|: || |: |   | ||  :         |||  :  |       | | || ||
Db      57 SVSEWLQSVQATMILSVIFSVLALFL--------FFCQLFTLTKGGRFYITGFFQILAGL 108

Qy     131 LGFIPVA--------WNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
                |        |::              :  :  | |  |  ::   :|::|||
Db     109 CVMSAAAIYTVRHSEWHV--------------NTDYSYGFAYILAWVAFPLALLSGII 152


RESULT  11
5194595-19
;Patent No. 5194595
;    APPLICANT: WATHEN, MICHAEL W.
;    TITLE OF INVENTION: CHIMERIC GLYCOPROTEINS CONTAINING
;IMMUNOGENIC SEGMENT OF THE GLYCOPROTEINS OF HUMAN RESPIRATORY
;SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES: 19
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/543,780
;      FILING DATE: 31-OCT-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 137,387
;      FILING DATE: 23-DEC-1987
;SEQ ID NO:19:
;      LENGTH: 681
5194595-19


  Query Match             7.2%;  Score 85;  DB 6;  Length 681;
  Best Local Similarity   20.1%;  Pred. No. 0.77;
  Matches   59;  Conservative   48;  Mismatches   92;  Indels   94;  Gaps   17;

Qy       1 MASLGLQLVGYI--LGLLGLLGTLVAMLLPSWK-------TSSYVGASIVTAVGFSKGLW 51
           |: :  ::: |:  | | |:: |       | ||       |::    | :      :| :
Db     289 MSIIKEEVLAYVVQLPLYGVIDT------PCWKLHTSPLCTTNTKEGSNICLTRTDRGWY 342

Qy      52 ME-------------CATHSTGITQCDIYSTLLGLPADIQAAQAMM-------------- 84
             :             |   |  :  ||  ::|  ||::|     :
Db     343 CDNAGSVSFFPQAETCKVQSNRVF-CDTMNSLT-LPSEINLCNVDIFNPKYDCKIMTSKT 400

Qy      85 -VTSSAISSLACIISVVG-MRCTVF-------------CQ--ESRAKDRVAVAGGVFFI- 126
            |:|| |:||  |:|  |  :||                |    ::  | |:|   ::::
Db     401 DVSSSVITSLGAIVSCYGKTKCTASNKNRGIIKTFSNGCDYVSNKGMDTVSVGNTLYYVN 460

Qy     127 -LGGLLGFI---PVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISS---------- 171
              |   ::   |:          :|| |||  | :|:      |||  |
Db     461 KQEGKSLYVKGEPII--------NFYDPLVFPSDEFD-----QLGISPSNPSEITSQITT 507

Qy     172 -LFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPP-KVKSEFN 222
             | |   |:      :: : ::      | ||    :  |  :|| |  ::|:
Db     508 ILASTTPGVKSTLQSTTVKTKNT--TTTQTQPSKPTTKQRQNKPPSKPNNDFH 558


RESULT  12
US-08-677-049-12
; Sequence 12, Application US/08677049
; Patent No. 5858707
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Guimaraes, M. Jorge
;    APPLICANT:  Bazan, J. Fernando
;    APPLICANT:  McClanahan, Terrill K.
;    APPLICANT:  Zlotnik, Albert
;    TITLE OF INVENTION:  PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
;    TITLE OF INVENTION:  NUCLEIC ACIDS; ANTIBODIES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/677,049
;      FILING DATE:  03-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/000,788
;      FILING DATE:  03-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0511
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-852-9196
;      TELEFAX:  415-496-1200
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  580 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  207..236
;      OTHER INFORMATION:  /note= "Encompasses TM 4 of Figure
;      OTHER INFORMATION:  4"
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  437..499
;      OTHER INFORMATION:  /note=" "Encompasses TM 9 and TM 10
;      OTHER INFORMATION:  of Figure 4"
US-08-677-049-12


  Query Match             7.2%;  Score 84.5;  DB 2;  Length 580;
  Best Local Similarity   25.2%;  Pred. No. 0.7;
  Matches   59;  Conservative   29;  Mismatches   93;  Indels   53;  Gaps   11;

Qy       8 LVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECAT------HSTGI 61
           ||    | : |||  :    |  :||  |||  :|: || | |  :  ||      :|||
Db     110 LVSTSLIVSGLLSAVQMFRLHVYKTRYYVGTGLVSVVGTSFGT-ITVATGTFNQMYSTGY 168

Qy      62 TQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAG 121
             |    | |   :   |:: ||   | |   :| :  |          :|     | |
Db     169 CPVDGSGNRLPCP---KGYGALLATSCLCSLLEIGLSFMSSRLL------KALFPPIVTG 219

Qy     122 GVFFILG-GLLGFIPVAW----------NLHGIL---RDFYSPLVPDSMKF-EIGEALYL 166
             |::|  |:|     |            :| |    |   ||   | :|  :|  ::
Db     220 PTVFLIGASLIGNAMKDWAGGSGTCSSNPGNGALCPSADAPHPLPWGSAEFIGLGFLVFA 279

Qy     167 GII------------SSLFSLIAGIILCFSCSSQRNRSNYYD--AYQAQPLAT 205
             ||            : :  |: | |:  :|       |:|     | |:|:
Db     280 TIILCERFGSPIMKSCAVIVGLLVGCIVAAAC-------GYFDRSGIDAAPVAS 326
```

```
RESULT  13
US-08-406-855A-19
; Sequence 19, Application US/08406855A
; Patent No. 5861309
;  GENERAL INFORMATION:
;    APPLICANT:  Jonathan A. Bard et al.
;    TITLE OF INVENTION:  DNA Encoding Human Alpha 1 Adrenergic
;    TITLE OF INVENTION:  Receptors and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/406,855A
;      FILING DATE:  21-AUG-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  41337-A-PCT-US/JPW/KDB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  521 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-406-855A-19


  Query Match             7.2%;  Score 84;  DB 2;  Length 521;
  Best Local Similarity   20.5%;  Pred. No. 0.68;
  Matches   58;  Conservative   36;  Mismatches   79;  Indels  110;  Gaps   14;

Qy      18 LLGTLVAMLLPSWKTSSY----------------------VGASIVTAVGFSKGLWM--- 52
             |   ::|  |  |:| :|                      ||  :|:| |   |:::
Db       5 LRSVMMAGYLSEWRTPTYRSTEMVQRLRMEAVQHSTSTAAVGGLVVSAQGVGVGVFLAAF 64

Qy      53 ---------------ECATHSTGITQ--------CDIYSTLLGLP--------------- 74
                          |  |   :|          |:  :   ||
Db      65 ILTAVAGNLLVILSVACNRHLQTVTNYFIVNLAVADLLLSAAVLPFSATMEVLGFWAFGR 124

Qy      75 --ADIQAAQAMMVTSSAISSLACIISV---VGMRCTV----FCQESRAKDRVAVAGGVFF 125
             |: ||  ::  :::| || | |||   ||:| ::       | :|   :|:   |
Db     125 TFCDVWAAVDVLCCTASILSL-CTISVDRYVGVRHSLKYPAIMTERKAAAILALLWAVAL 183

Qy     126 I--LGGLLGFIPVAWNLHGILRDFYSPLVPD----SMKFEIGEALYLGIISSLFSLIAGI 179
           :  :| |||      |           |: ||      :  |:| |::  : |  | |  :
Db     184 VVSVGPLLG-----WK---------EPVPPDERFCGITEEVGYAIFSSVCS--FYLPMAV 227

Qy     180 ILCFSCSSQRNRSNYYDAYQAQPLATRS-----SPRPGQPPKV 217
           |:  |           |     |||        ||:  :|
Db     228 IVVMYCR----------VYVVARSTTRSLEAGIKREPGKASEV 260


RESULT  14
US-09-206-899-19
; Sequence 19, Application US/09206899
; Patent No. 6083705
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Jonathan A. Bard et al.
;    TITLE OF INVENTION:  DNA Encoding Human Alpha 1 Adrenergic
;    TITLE OF INVENTION:  Receptors and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/206,899
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/406,855
;      FILING DATE:  21-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  41337-A-PCT-US/JPW/KDB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  521 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-206-899-19


  Query Match             7.2%;  Score 84;  DB 3;  Length 521;
  Best Local Similarity   20.5%;  Pred. No. 0.68;
  Matches   58;  Conservative   36;  Mismatches   79;  Indels  110;  Gaps   14;

Qy      18 LLGTLVAMLLPSWKTSSY----------------------VGASIVTAVGFSKGLWM--- 52
           |   ::|  |  |:| :|                      ||  :|:| |   |:::
Db       5 LRSVMMAGYLSEWRTPTYRSTEMVQRLRMEAVQHSTSTAAVGGLVVSAQGVGVGVFLAAF 64

Qy      53 ---------------ECATHSTGITQ--------CDIYSTLLGLP--------------- 74
                          |  |   :|          |:  :   ||
Db      65 ILTAVAGNLLVILSVACNRHLQTVTNYFIVNLAVADLLLSAAVLPFSATMEVLGFWAFGR 124

Qy      75 --ADIQAAQAMMVTSSAISSLACIISV---VGMRCTV----FCQESRAKDRVAVAGGVFF 125
             |: ||  ::  :::| || | |||   ||:| ::       | :|   :|:   |
Db     125 TFCDVWAAVDVLCCTASILSL-CTISVDRYVGVRHSLKYPAIMTERKAAAILALLWAVAL 183

Qy     126 I--LGGLLGFIPVAWNLHGILRDFYSPLVPD----SMKFEIGEALYLGIISSLFSLIAGI 179
           :  :| |||     |           |: ||     :  |:| |::  : |  | |   :
Db     184 VVSVGPLLG-----WK---------EPVPPDERFCGITEEVGYAIFSSVCS--FYLPMAV 227

Qy     180 ILCFSCSSQRNRSNYYDAYQAQPLATRS-----SPRPGQPPKV 217
           |:   |             |      |||        ||:  :|
Db     228 IVVMYCR----------VYVVARSTTRSLEAGIKREPGKASEV 260


RESULT  15
US-08-518-474-8
; Sequence 8, Application US/08518474
; Patent No. 5599920
;  GENERAL INFORMATION:
;    APPLICANT:  Patel, Pragna I.
;    APPLICANT:  Suter, Ueli
;    APPLICANT:  Snipes, G. J.
;    APPLICANT:  Welcher, Andrew
;    APPLICANT:  DeLeon, Marino
;    APPLICANT:  Lupski, James R.
;    APPLICANT:  Shooter, Eric M.
;    TITLE OF INVENTION:  PERIPHERAL MYELIN PROTEIN CODING
;    TITLE OF INVENTION:  SEQUENCE AND METHOD
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fulbright & Jaworski
;      STREET:  1301 McKinney, Suite 5100
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77010-3095
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/518,474
;      FILING DATE:  21-AUG-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/308,488
;      FILING DATE:  19-SEP-1994
;      APPLICATION NUMBER:  US 07/879,623
;      FILING DATE:  06-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul, Thomas D.
;      REGISTRATION NUMBER:  32,714
;      REFERENCE/DOCKET NUMBER:  D-5331-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713/651-5151
;      TELEFAX:  713/651-5246
;      TELEX:  762829
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  144 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-518-474-8


  Query Match             7.1%;  Score 83;  DB 1;  Length 144;
  Best Local Similarity   29.9%;  Pred. No. 0.14;
  Matches   35;  Conservative   17;  Mismatches   47;  Indels   18;  Gaps    6;

Qy      12 ILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTLL 71
           :| |||:|   :|:|:  :  |: |   :|   | :  ||  | | : |  |   ||:
Db       2 LLLLLGILFLHIAVLVLLF-VSTIVSQWLV-GNGHTTDLWQNCTTSALGAVQ-HCYSS-- 56

Qy      72 GLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGGVFFILG 128
              :   :|: || |: |   | ||  :          |||           || |:| |
Db      57 SVSEWLQSVQATMILSVIFSVLALFL--------FFCQLF-----TLTKGGRFYITG 100


Search completed: June  2, 2002, 20:13:56
Job time: 4496 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:58:20 ; Search time 93.15 Seconds
(without alignments)
237.258 Million cell updates/sec

Title: US-09-663-600A-186
Perfect score: 1174
Sequence: 1 MASLGLQLVGYILGLLGLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	423.5	36.1	280	2	A39484	androgen-withdrawa
2	100	8.5	400	2	D97643	probable efflux pr
3	100	8.5	400	2	AG2866	MFS permease [drug
4	99.5	8.5	330	2	C69648	2-keto-3-deoxygluc
5	99.5	8.5	463	2	F90285	metabolite transpo
6	97.5	8.3	484	2	C75609	amino acid ABC tra
7	95.5	8.1	458	2	H71657	NADH dehydrogenase
8	93	7.9	523	2	T11916	NADH dehydrogenase
9	91	7.8	160	2	JN0503	peripheral myelin
10	90.5	7.7	487	2	T19237	hypothetical prote
11	90.5	7.7	617	2	F71019	hypothetical prote
12	88.5	7.5	429	2	A97241	permease [imported
13	88.5	7.5	521	2	A99549	amino acid permeas
14	88.5	7.5	593	2	H71283	conserved hypothet
15	88	7.5	160	2	A41144	growth arrest-rela
16	88	7.5	421	2	A99309	membrane transport
17	88	7.5	906	2	G90281	conserved hypothet
18	87.5	7.5	268	2	A71086	hypothetical prote
19	87.5	7.5	420	2	H96534	probable Na+/H+ an
20	87.5	7.5	519	2	T15364	hypothetical prote
21	87.5	7.5	690	2	S35251	probable membrane
22	87	7.4	157	2	G02355	tumor-associated m
23	87	7.4	160	2	S21721	growth arrest-spec
24	87	7.4	223	2	E87654	hypothetical prote
25	87	7.4	318	2	AG3576	iron(III) dicitrat
26	87	7.4	414	2	F65097	hypothetical 43.5
27	87	7.4	414	2	C91125	probable transport
28	87	7.4	414	2	B85970	probable transport
29	86.5	7.4	255	2	E64604	conserved hypothet
30	86.5	7.4	402	2	AI1853	sodium-dependent n
31	86.5	7.4	432	2	A70714	hypothetical prote
32	86.5	7.4	458	2	D97799	NADH dehydrogenase
33	86.5	7.4	497	2	AD3525	amino acid permeas
34	86.5	7.4	547	2	B64963	membrane protein y
35	86	7.3	398	2	JQ0113	2-keto-3-deoxygluc
36	86	7.3	409	2	S29124	membrane glycoprot
37	86	7.3	562	2	T52481	cytochrome-c oxida
38	85.5	7.3	216	2	I38474	olfactory receptor
39	85.5	7.3	371	2	D75266	cell division prot
40	85.5	7.3	414	2	AG0894	probable membrane
41	85.5	7.3	445	2	T11097	NADH dehydrogenase
42	85	7.2	233	2	T31865	hypothetical prote
43	85	7.2	502	2	B70845	probable sygar tra
44	85	7.2	531	2	T11074	NADH dehydrogenase
45	84.5	7.2	454	2	C86766	hypothetical prote

ALIGNMENTS

RESULT 1
A39484
androgen-withdrawal apoptosis protein RVP1, prostatic - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 04-Mar-2000
C;Accession: A39484
R;Briehl, M.M.; Miesfeld, R.L.
Mol. Endocrinol. 5, 1381-1388, 1991
A;Title: Isolation and characterization of transcripts induced by androgen withdrawal
A;Reference number: A39484; MUID:92130987
A;Accession: A39484
A;Molecule type: mRNA
A;Residues: 1-280 <BRI>
A;Cross-references: GB:M74067; NID:g205857; PIDN:AAA41760.1; PID:g205858
C;Genetics:
A;Gene: RVP.1
C;Superfamily: rat androgen-withdrawal apoptosis protein RVP1

Query Match 36.1%; Score 423.5; DB 2; Length 280;
Best Local Similarity 37.7%; Pred. No. 4.2e-29;
Matches 84; Conservative 50; Mismatches 74; Indels 15; Gaps 3;

```
Qy       3 SLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGIT 62
           |: |:: |  | :|| | |:|   || |: |:::|:||:||    :|||| |   |||
Db       2 SMSLEITGTSLAVLGWLCTIVCCALPMWRVSAFIGSSIITAQITWEGLWMNC-VQSTGQM 60

Qy      63 QCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGG 122
           || :| :|| || |:|||:|::| |  :::   ::::|| :||   |:  || :: :  |
Db      61 QCKMYDSLLALPQDLQAARALIVVSILLAAFGLLVALVGAQCTNCVQDETAKAKITIVAG 120

Qy     123 VFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILC 182
           | |:|  :|  :||:|: : |:||||:||||:: | |:|  ||:|  ::   |: | :||
Db     121 VLFLLAAVLTLVPVSWSANTIIRDFYNPLVPEAQKREMGTGLYVGWAAAALQLLGGALLC 180

Qy     183 FSCSSQRNRSNYYDAYQAQPLAT------------RSSPRPG 212
           ||  :   :     | | | :|             :| |||
Db     181 CSCPPREKYAPTKILYSA-PRSTGPGTGTGTAYDRKTTSERPG 222
```

RESULT 2
D97643
probable efflux protein cj1687 [imported] - Agrobacterium tumefaciens (strain C58, Ce
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: D97643
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A;Reference number: A97359; PMID:11743194
A;Accession: D97643

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-400 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88101.1; PID:g15157533; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4286
A;Map position: circular chromosome


  Query Match             8.5%;  Score 100;  DB 2;  Length 400;
  Best Local Similarity   21.2%;  Pred. No. 0.41;
  Matches   44;  Conservative   31;  Mismatches   67;  Indels   66;  Gaps    7;

Qy       8 LVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIY 67
           : || :|:  | | |:|  |   :  : :| ::: | |          :||
Db     243 VAGYAIGIFNLAGALMA--LSFGRLRARIGVTVILAAGL--------VLMATGF------ 286

Qy      68 STLLGLPADIQAAQAMMVTSSAISSLACIISVVG----------MRCTVFCQESRAKDRV 117
                     |:: |:  :||:   ::|||          |  |:     | : |:
Db     287 --------------ALLATAGGLSSMLSALAVVGAGLGAVMPSIMSTTIMLAPLRLRGRI 332

Qy     118 A--VAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSL 175
           |  |   :|                |  | ||||        | |     |: :|::
Db     333 AGIVTASMF------------------LGHFISPLVSQPWIARFGFATTYRDIALVFAV 373

Qy     176 IAGIILCFSCSSQRNRSNYYDAYQAQPL 203
           :||:    :   :|         :||||
Db     374 MAGLAAIATIFQRRT-----GIRKAQPL 396


RESULT   3
AG2866
MFS permease [drug] [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AG2866
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AG2866
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-400 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43349.1; PID:g17740844; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2361
A;Map position: circular chromosome


  Query Match             8.5%;  Score 100;  DB 2;  Length 400;
  Best Local Similarity   21.2%;  Pred. No. 0.41;
  Matches   44;  Conservative   31;  Mismatches   67;  Indels   66;  Gaps    7;

Qy       8 LVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIY 67
           : || :|:  | | |:|  |   :  : :| ::: | |          :||
Db     243 VAGYAIGIFNLAGALMA--LSFGRLRARIGVTVILAAGL--------VLMATGF------ 286

Qy      68 STLLGLPADIQAAQAMMVTSSAISSLACIISVVG----------MRCTVFCQESRAKDRV 117
                     |:: |:  :||:   ::|||          |  |:     | : |:
Db     287 --------------ALLATAGGLSSMLSALAVVGAGLGAVMPSIMSTTIMLAPLRLRGRI 332

Qy     118 A--VAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSL 175
           |  |   :|                |  | ||||        | |     |: :|::
Db     333 AGIVTASMF------------------LGHFISPLVSQPWIARFGFATTYRDIALVFAV 373

Qy     176 IAGIILCFSCSSQRNRSNYYDAYQAQPL 203
           :||:    :   :|         :||||
Db     374 MAGLAAIATIFQRRT-----GIRKAQPL 396


RESULT   4
C69648
2-keto-3-deoxygluconate permease kdgT - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: C69648
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: C69648
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-330 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14126.1; PID:g26346
A;Experimental source: strain 168
C;Genetics:
A;Gene: kdgT


  Query Match             8.5%;  Score 99.5;  DB 2;  Length 330;
  Best Local Similarity   24.4%;  Pred. No. 0.37;
  Matches   51;  Conservative   30;  Mismatches   77;  Indels   51;  Gaps   11;

Qy       2 ASLGLQLVGYI----------LGLLGLLGTLVAMLLP--SWKTSSYVGASIVTAVGF-- 46
           |::  :  |||           :|   ||| : |  :|    ::  : | |::  |
Db      60 ATIDFRSSGYIARKGITLLLGKIGFAALLGVIAAQFIPDDGIQSGFFAGLSVLAIVAVMN 119

Qy      47 --SKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRC 104
             : ||::    |             :|   |   | | : | |          | |:
Db     120 ETNGGLYLALMNH------------MGRKED-AGAFAFISTESGPFMTMVTFGVTGL-- 163

Qy     105 TVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVP---DSMKFEIG 161
             |  |: |      |  : |:|| :||      ||   ||| :| :||       | :|
Db     164 AAFPWETLA------ATVIPFLLGCILG------NLDHDLRDLFSKVVPAIIPFFAFSLG 211

Qy     162 EALYLGII--SSLFSLIAGI-ILCFSCSS 187
             |  |::  | |  :  |: ::  | ||
Db     212 NTLNFGMLIQSGLLGIFIGVSVVILSGSS 240


RESULT   5
F90285
metabolite transport related protein SSO1305 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: F90285
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: F90285
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-463 <KUR>
A;Cross-references: GB:AE006641; NID:g13814507; PIDN:AAK41541.1; GSPDB:GN00155
```

```
C;Genetics:
A;Gene: SSO1305


  Query Match             8.5%;  Score 99.5;  DB 2;  Length 463;
  Best Local Similarity   23.8%;  Pred. No. 0.53;
  Matches   44;  Conservative   35;  Mismatches   59;  Indels   47;  Gaps   10;

Qy       6 LQLVGYILGLLGLLGLTLVAMLLPSWKTSSYVGASIVTAVGFS-------KGLWMECAT-- 56
           :|:  ::  ::|::|  : :      :   |  :|:: |:||:       | ||:
Db     288 VQIAWFVTAIIGVVGVYMQLSDKVNRKLLYAVSSLICAIGFALPIGLPFKILWVALFNVF 347

Qy      57 -----HSTGI-TQCDIYSTLLGLPADIQ-AAQ---------AMMVTSSAISSLACIISVV 100
                |   :  |  ::|| |  | :|:  ||          |: : |  : |   ||:|:
Db     348 LFGFGHGMALWPQTRVWSTEL-FPTEIRNTAQGFVWGWMRIALGIWSIFVPS---IITVI 403

Qy     101 GMRCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHG-----ILRDFYSPLVPDS 155
           |              :|     ||||  :|| : |  : ||     :::|||   || |
Db     404 GY------------SAIAAVATSFFILNIILGLL-VGPDTHGKSLEVVIKDFYGGKVPVS 450

Qy     156 MKFEI 160
             |:
Db     451 KVVEV 455


RESULT   6
C75609
amino acid ABC transporter, permease protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: C75609
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: C75609
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-484 <WHI>
A;Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12224.1; PID:g646051
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0136
A;Map position: 2


  Query Match             8.3%;  Score 97.5;  DB 2;  Length 484;
  Best Local Similarity   26.1%;  Pred. No. 0.82;
  Matches   55;  Conservative   25;  Mismatches   84;  Indels   47;  Gaps    9;

Qy      14 GLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTLLGL 73
           |: | |   |  ||  | |   |       |  |    | :  ||:       : : ||||
Db     164 GVAGALAFAVPPLLGKWNTLGLVQEYRNVADTF----WPQLGTHAALSVIALLLAALLGL 219

Qy      74 PADIQAAQAMMVTSSAI---SSLACIISVV--GMRCTVFCQESRAKD------------ 115
           |  | ||:   :    :   | |  | ||   |:   ||    |
Db     220 PLGIAAARNTRLAGGVLGFASFLQTIPSVALFGLLLPVFSALGRGVSVGAFLAWSGAALL 279

Qy     116 ---------RVAVAGGVFFILG----------GLLGFIPVAW-NLHGILRDFYS---PLV 152
                   |:|: ||:  :||          |||  :  ||    |: || :|   ||
Db     280 LGLALTRVRRLALPGGLLALLGAQALLLLAGLGLLQVLQGAWFGGDGLSRDAFSLSAPLA 339

Qy     153 PDSMKFEIGEALYLGIISSLFSLIAGIILCF 183
           ::  || |  |  : :|::|:  ::  |
Db     340 SWGVR-GIGAAPALFAL-TLYALLPIVVNTF 368


RESULT   7
H71657
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain N RP537 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C;Accession: H71657
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499
A;Accession: H71657
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-458 <AND>
A;Cross-references: GB:AJ235272; GB:AJ235269; NID:g3861033; PIDN:CAA14986.1; PID:g386
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: nuoN1; RP537
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 2
C;Keywords: membrane-associated complex; NAD; oxidoreductase


  Query Match             8.1%;  Score 95.5;  DB 2;  Length 458;
  Best Local Similarity   23.6%;  Pred. No. 1.2;
  Matches   59;  Conservative   33;  Mismatches   73;  Indels   85;  Gaps   12;

Qy      13 LGLLGLLGTLVAMLLP-------------------SWKTSSYVGASIVTAVGFSKGLWM 52
           | |: |||   |:::|                     ::| |||            :|:|
Db      10 LTLIALLGQCFALMIPNKNRIIYNIVILLCIISIFLTFKYSSY-----------EGIWH 57

Qy      53 ECAT-HSTGITQCDI------------YSTLLGLPADIQAAQAMMVT----SSAISSLA 94
             ||  : ||::  |              || |:|     :    |:::     ||||
Db      58 SFATERNIGISKSIILLFTIVSLIIYRDYSILVGETLKFEFITLMLLSIVGIFVAISSRN 117

Qy      95 CIISVVGMRCTVFCQESRA----KDRVAVAGGV-FFILGGLLGFIPVAWNLHGI------ 143
             ::   ||  |     : |     |  :  | : :|||| |:  :    :| ||
Db     118 FLLLFCGMELTALTSYALAGFKLNDIKSSEGALKYFILGSLVSCL----SLFGISFIYGF 173

Qy     144 -----LRDFYSPLVPDSMKFEIGEALYLGII---SSLFSLIAGIILCFSCSSQRNRSNYY 195
                |    |  ||   ||   | :||:   ||:|  :|   | |
Db     174 GGSIQFDDILHQLNNDS---EIKPGLIIGIVLFLSSIFFKLASSPLHFWIP--------- 221

Qy     196 DAYQAQPLAT 205
           | |:  |:::
Db     222 DVYEGSPISS 231


RESULT   8
T11916
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Prototheca wickerhamii mitocho
C;Species: mitochondrion Prototheca wickerhamii
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 12-Nov-1999
C;Accession: T11916
R;Wolff, G.; Plante, I.; Lang, B.F.; Kueck, U.; Burger, G.
J. Mol. Biol. 237, 75-86, 1994
A;Title: Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca
A;Reference number: Z17373; MUID:94180393
A;Accession: T11916
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-523 <WOL>
A;Cross-references: EMBL:U02970; NID:g467843; PID:g467848; PIDN:AAD12635.1
A;Experimental source: strain HB-8
C;Genetics:
A;Genome: mitochondrion
A;Note: nad4
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylatio


  Query Match             7.9%;  Score 93;  DB 2;  Length 523;
  Best Local Similarity   24.2%;  Pred. No. 2.2;
  Matches   45;  Conservative   37;  Mismatches   50;  Indels   54;  Gaps   11;

Qy       8 LVGYILGLLG--LLGTLVAMLLPSWKTSSY----VGASIVTAVGFSKGLWMECATHST-- 59
           :: ||  :|   |||  :  : :|||||  :     : :|::| :   |  ||:|  : |
Db      24 IMSYIEMVLAIPLLGAIALLFVPSWKTQTIRNIALNSSLLTFL-ISLLLWIEFDSSSALF 82
```

```
Qy      60 ----GITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKD 115
               |:   ::||       |:  |:|   :||: |:|   : | |:
Db      83 QFTDGVCSPNVYS-------DVTLAKA--ASSSSFSALNFALGVDGI------------ 120

Qy     116 RVAVAGGVFFILGGLLGFIP----VAW-NLHGILRDFYSPLVPDSMKFEIGEALYLGIIS 170
                 :|||:   |  :|    |:| |:   ::::        : | : | | | : |
Db     121 ------SLFFIILTTL-LVPICILVSWNNIEVYVKEY-------CIAFLVLETLMLTVFS 166

Qy     171 SLFSLI 176
             |  |:
Db     167 VLDLLL 172


RESULT   9
JN0503
peripheral myelin protein 22 - human
N;Alternate names: Charcot-Marie-Tooth; GAS-3 protein; growth arrest-specific protein 3;
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Jun-2000
C;Accession: JN0503; JC1190; A56697; S25537
R;Edomi, P.; Martinotti, A.; Colombo, M.P.; Schneider, C.
Gene 126, 289-290, 1993
A;Title: Sequence of human GAS3/PMP22 full-length cDNA.
A;Reference number: JN0503; MUID:93246261
A;Accession: JN0503
A;Molecule type: mRNA
A;Residues: 1-160 <EDO>
A;Cross-references: GB:L03203; NID:g182984; PIDN:AAA58495.1; PID:g182985
R;Hayasaka, K.; Himoro, M.; Nanao, K.; Sato, W.; Miura, M.; Uyemura, K.; Takahashi, E.;
Biochem. Biophys. Res. Commun. 186, 827-831, 1992
A;Title: Isolation and sequence determination of cDNA encoding PMP-22 (PAS-II/SR13/GAS-3
A;Reference number: JC1190; MUID:92360032
A;Accession: JC1190
A;Molecule type: mRNA
A;Residues: 1-160 <HAY>
A;Cross-references: GB:D11428; NID:g220009; PIDN:BAA01995.1; PID:g220010
A;Experimental source: fetus spinal cord
R;Valentijn, L.J.; Baas, F.; Wolterman, R.A.; Hoogendijk, J.E.; van den Bosch, N.H.A.; Z
Nature Genet. 2, 288-291, 1992
A;Title: Identical point mutations of PMP-22 in Trembler-J mouse and Charcot-Marie-Tooth
A;Reference number: A56697; MUID:93265161
A;Accession: A56697
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-160 <VAL>
A;Cross-references: GB:S61788
R;Colombo, M.
submitted to the EMBL Data Library, May 1992
A;Reference number: S25537
A;Accession: S25537
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 61-160 <COL>
A;Cross-references: EMBL:X65968; NID:g31652; PIDN:CAA46781.1; PID:g31653
C;Genetics:
A;Gene: GDB:PMP22; GAS3
A;Cross-references: GDB:134190
A;Map position: 17p12-17p11.2
C;Superfamily: growth arrest-specific protein
C;Keywords: myelin; transmembrane protein


  Query Match             7.8%;  Score 91;  DB 2;  Length 160;
  Best Local Similarity   21.7%;  Pred. No. 0.96;
  Matches   41;  Conservative   33;  Mismatches   69;  Indels   46;  Gaps    8;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           |  | | ::   : :| ||   |: ::  |          :   | :  ||  |:| |:|
Db       1 MLLLLLSIIVLHVAVLVLL--FVSTIVSQW----------IVGNGHATDLWQNCSTSSSG 48

Qy      61 ITQCDIYSTLLGLPAD-IQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAV 119
               :::      | : :|: || |: |   |    |:|:    | :|      :  :
```

```
Db      49 ----NVHHCFSSSPNEWLQSVQATMILSIIFS----ILSLFLFFCQLFTLTKGGRFYIT- 99

Qy     120 AGGVFFILGGLLGFIPVA--------WNLHGILRDFYSPLVPDSMKFEIGEALYLGIISS 171
             |:| || ||       |        |:|:                :  | |  |  ::
Db     100 --GIFQILAGLCVMSAAAIYTVRHPEWHLNS-------------DYSYGFAYILAWVAF 143

Qy     172 LFSLIAGII 180
             :|::|:|
Db     144 PLALLSGVI 152


RESULT  10
T19237
hypothetical protein C27A7.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T19237; T19502
R;Harris, B.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19095
A;Accession: T19237
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-487 <WIL>
A;Cross-references: EMBL:Z92825; PIDN:CAB07317.1; GSPDB:GN00023; CESP:C27A7.6
A;Experimental source: clone C13C4
R;Harris, B.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19132
A;Accession: T19502
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-487 <WI2>
A;Cross-references: EMBL:Z81041; PIDN:CAB02792.1; GSPDB:GN00023; CESP:C27A7.6
A;Experimental source: clone C27A7
C;Genetics:
A;Gene: CESP:C27A7.6
A;Map position: 5
A;Introns: 30/3; 51/3; 87/3; 109/3; 183/1; 215/3; 264/2; 302/2; 330/3; 349/3; 372/1;


  Query Match             7.7%;  Score 90.5;  DB 2;  Length 487;
  Best Local Similarity   28.7%;  Pred. No. 3.3;
  Matches   51;  Conservative   17;  Mismatches   79;  Indels   31;  Gaps    7;

Qy       5 GLQLVGYILGLLGLLGTLVAMLLP--SWKTSSYVGASIVTAVGFSKG---LWMECATHST 59
           | :: ||   :  ::||| ::|    : ||  :     |  ||||     | |      |
Db     297 GYEMAGYPTAVCAIVGTLTSLLAGHIADKTRKFKEIIRVCTVGFSCSVITLRMFLNQPRT 356

Qy      60 GITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISV-VGMRCTVFCQESRAKDRVA 118
           |:    |  || |                | |     | | :|:  |    |       |
Db     357 GLFDSIIVYTLCG-------------CLGAFSIPQFPIGVELGVETTFPVME-------A 396

Qy     119 VAGGVFFILGGLLGF-IPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSL 175
             : ||  | | |  | || | |    |  ||:     | || :     | |:| : ||
Db     397 TSSGVLVIFGSLFMFIIPFAQNYTEKLHLFYA----QSWKFALDVTCGLSIVSVILSL 450


RESULT  11
F71019
hypothetical protein PH1451 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: F71019
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137
A;Accession: F71019
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-617 <KAW>
```

```
A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30558.1; PID:g3257875
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1451
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5


  Query Match             7.7%;  Score 90.5;  DB 2;  Length 617;
  Best Local Similarity   24.4%;  Pred. No. 4.2;
  Matches   52;  Conservative   33;  Mismatches   81;  Indels   47;  Gaps    9;

Qy       5 GLQLVGYILGLLGLL-----GTLVAMLLPSWKTSSY----------VGASIVTAVGFS-- 47
           |  | |||:  || |     | | |:     |  :|          ||  | |::|
Db     273 GTPLFGYIIAFLGGLTIIIGGVLAALQEDIRKLFAYSSISQVGYILVGLGIGTSLGIEAA 332

Qy      48 ----------KGL-WMECAT--HSTGITQCDIYSTLLGLPADIQAAQAMMVTSS----AI 90
                    ||| ::  ||  : || |:   :  |         |: |  | :    ||
Db     333 IYHAISHALFKGLFFLVVATIIYRTGKTEFKDFGGL---------AEKMPFTFAMAFIAI 383

Qy      91 SSLACIISVVGMRCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSP 150
             ||| |  :||        |:    :: : ||: |  |  :||: :    : :
Db     384 LSLAGIPPLVGFASKWLIFEAVISQKLPILGGMLF-FGSAIGFVYLIRFTYAVWFGQRPS 442

Qy     151 LVPDSMKFEIGEALYLGIISSL---FSLIAGII 180
             : |:    :  |: :||: :|   | :  |::
Db     443 DIEDTKDAPLPLAIGMGILGTLNVVFGVAPGLV 475


RESULT  12
A97241
permease [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: A97241
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A97241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-429 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80716.1; PID:g15025810; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2772
C;Superfamily: conserved hypothetical protein HI0125


  Query Match             7.5%;  Score 88.5;  DB 2;  Length 429;
  Best Local Similarity   22.8%;  Pred. No. 4.3;
  Matches   41;  Conservative   38;  Mismatches   76;  Indels   25;  Gaps    8;

Qy      13 LGLLGLLG-TLVAMLLPSWKTSSYVGASIVTAVGF-------SKGLWMECATH------S 58
           :||  |   |:   |  ||||:  : ||::  : |        : | ::          |
Db      79 MGLNALFTYTICIQLHFSWKTA--LAASLIEGIIFLVLNVFKIRQLILDSVPQTLKYAIS 136

Qy      59 TGITQCDIYSTLLGL--PADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDR 116
             ||      : | :||     |  ::  :|| ::: | | :::|:|:        |
Db     137 IGI---GFFITFIGLQDAGIIVGSKGTLVTLASLKSPAVLLAVIGVLLIAVLYNKNIKGS 193

Qy     117 VAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLI 176
             :   | ::|| : |   ||    ||:  |   : |  ::|:   |  :|||:  : :::
Db     194 FVIGMLVIYVLGIIFG---VAKAPSGIV-SFPPSVAPVFLQFDFKSAAVIGIVPVILTML 249


RESULT  13
A99549
amino acid permease [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: A99549
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: A99549
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-521 <KUR>
A;Cross-references: GB:AL445566; PID:g14089711; PIDN:CAC13470.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_2970
A;Genetic code: SGC3


  Query Match             7.5%;  Score 88.5;  DB 2;  Length 521;
  Best Local Similarity   21.2%;  Pred. No. 5.3;
  Matches   50;  Conservative   38;  Mismatches   63;  Indels   85;  Gaps   11;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLL--------------PSWKTSSYVGA--SIVTA 43
             |:    : ||: : |  |||::| : :                ||  :   |   |:  |
Db       1 MSEKTFKKVGFFVALSMLLGSVVGIGIFFKNGSVGRAVEHNGTSWLLAWVFGGIISLAAA 60

Qy      44 VGFSK-----------GLWMECATHSTG-------------ITQCDIYSTLLG------ 72
           : ||:            | |    :|  |             :   |   :||
Db      61 INFSEISFLKNTKIAGIGNW----SHKVGDKRFGYFVSISFTLFYSGIIQIILGFFTAEI 116

Qy      73 -------LPADIQAAQ-AMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGGVF 124
                | |:|:      :|  : || :  |:::| ::                | |||
Db     117 FFHWLNLLGANIKMEMWHSLVLGTVISFIFIILNIVSIK---------------ASGVF 160

Qy     125 FILGGLLGFIP-VAWNLHGI-------LRDFYSPLVPDSMKFEIGEALYLGIISSL 172
             ::  :| |:| :|  | |:       | |  :  ||:  | :|||  :  :|::|
Db     161 QVITTILKFLPLIATILVGLIFVSTHTLEDGANAFVPNKTK-KIGEFSFANVIAAL 215


RESULT  14
H71283
conserved hypothetical integral membrane protein TP0771 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C;Accession: H71283
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770
A;Accession: H71283
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-593 <COL>
A;Cross-references: GB:AE001248; GB:AE000520; NID:g3323074; PIDN:AAC65739.1; PID:g332
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0771


  Query Match             7.5%;  Score 88.5;  DB 2;  Length 593;
  Best Local Similarity   23.2%;  Pred. No. 6;
  Matches   45;  Conservative   36;  Mismatches   72;  Indels   41;  Gaps    8;

Qy       4 LGLQLVGYILGL----LGLLGTLVAMLLPSWKTSSYV-----GASIVTAVGFSKGLWMEC 54
             ||  ::|:  ||    |||| :|:  |  | :  |::       ::   ||   |  :
Db     177 LGESIMGF--GLLFTGLGLLSSLIPPL--SVEELSFLKIAVEDRALSVFVGLLSGFVLTV 232

Qy      55 ATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAK 114
             ||:  |   : :   |    :: | | :: |:  |::   |: :|          |:
Db     233 ILHSSSATTAIVLTMAFGGVIGVEFAAASVLGSNVGSTIDAAIAAIG---------SKLN 283

Qy     115 DRVAVAGGVFFILGGLLGFIPVAWNLHGILRDFYSP-------LVPDSMKFEIGEALYLG 167
           | | |  | | : | | |:            |: |       | | :  |:    : |
```

```
Db    284 ARRAAAVHVLFNVFGALVFLM----------FFHPVLALLCVLTPKNSGFD-NITVRLA 331

Qy    168 IISSLFSLIAGIIL 181
          :  |:|:::  ||:
Db    332 LFHSMFNIVNTIIV 345
```

RESULT 15
A41144
growth arrest-related myelin protein precursor, sciatic nerve - rat
N;Alternate names: Schwann cell membrane glycoprotein, SAG
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 29-Sep-1999
C;Accession: A41144; S18550; A44826
R;Welcher, A.A.; Suter, U.; De Leon, M.; Snipes, G.J.; Shooter, E.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 7195-7199, 1991
A;Title: A myelin protein is encoded by the homologue of a growth arrest-specific gene.
A;Reference number: A41144; MUID:91334432
A;Accession: A41144
A;Molecule type: mRNA
A;Residues: 1-160 <WEL>
A;Cross-references: GB:M69139; NID:g207063; PIDN:AAA73063.1; PID:g207064
R;Spreyer, P.; Kuhn, G.; Hanemann, C.O.; Gillen, C.; Schaal, H.; Kuhn, R.; Lemke, G.; Mu
EMBO J. 10, 3661-3668, 1991
A;Title: Axon-regulated expression of a Schwann cell transcript that is homologous to a
A;Reference number: S18550; MUID:92037527
A;Accession: S18550
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-29,'E',31-160 <SPR>
A;Cross-references: EMBL:X62431; NID:g55903; PIDN:CAA44297.1; PID:g55904
R;Dieperink, M.E.; O'Neill, A.; Magnoni, G.; Wollmann, R.L.; Heinriksoń, R.L.; Zucher-Ne
J. Neurosci. 12, 2177-2185, 1992
A;Title: SAG: a Schwann cell membrane glycoprotein.
A;Reference number: A44826; MUID:92300442
A;Accession: A44826
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-21,'K',23-25,'X',27,'R',29-30 <DIE>
A;Note: sequence extracted from NCBI backbone (NCBIP:106581)
C;Comment: The predicted signal sequence may not be cleaved in some cell types.
C;Superfamily: growth arrest-specific protein
C;Keywords: glycoprotein; Schwann cell; transmembrane protein

  Query Match   7.5%; Score 88; DB 2; Length 160;
  Best Local Similarity   25.8%; Pred. No. 1.7;
  Matches   46; Conservative   26; Mismatches   70; Indels   36; Gaps   8;

```
Qy     12 ILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTLL 71
          :| |||:|   :|:|:  :  |: |   :|   |    ||  | | : |  |   ||:
Db      2 LLLLLGILFLHIAVLVLLF-VSTIVSQWLV-GNGHRTDLWQNCTTSALGAVQ-HCYSS-- 56

Qy     72 GLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQ-ESRAKDRVAVAGGVFFILGGL 130
            :   :|: || |: |   | |:  :         |||  :  |       ||| || ||
Db     57 SVSEWLQSVQATMILSVIFSVLSLFL--------FFCQLFTLTKGGRFYITGVFQILAGL 108

Qy    131 LGFIPVA--------WNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
                |        |:::               :  | |  |  ::   :|::|||
Db    109 CVMSAAAIYTVRHSEWHVNN-------------DYSYGFAYILAWVAFPLALLSGII 152
```

Search completed: June 2, 2002, 20:12:43
Job time: 4463 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 2, 2002, 20:09:12 ; Search time 65.66 Seconds
(without alignments)
135.630 Million cell updates/sec

Title: US-09-663-600A-186
Perfect score: 1174
Sequence: 1 MASLGLQLVGYILGLLGLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1174	100.0	230	1	CLD2_HUMAN	P57739 homo sapien
2	1119	95.3	230	1	CLD2_CANFA	Q95km6 canis famil
3	1096	93.4	230	1	CLD2_MOUSE	O88552 mus musculu
4	571	48.6	239	1	CLDE_HUMAN	O95500 homo sapien
5	561.5	47.8	239	1	CLDE_MOUSE	Q9z0s3 mus musculu
6	459.5	39.1	208	1	CLDY_BRARE	Q9yh91 brachydanio
7	450.5	38.4	218	1	CLD3_CANFA	Q95km5 canis famil
8	450	38.3	220	1	CLD3_HUMAN	O15551 homo sapien
9	446	38.0	210	1	CLD4_MOUSE	O35054 mus musculu
10	443.5	37.8	209	1	CLD4_CERAE	O19005 cercopithec
11	443	37.7	219	1	CLD6_MOUSE	Q9z262 mus musculu
12	442.5	37.7	209	1	CLD4_HUMAN	O14493 homo sapien
13	441	37.6	217	1	CLD9_HUMAN	O95484 homo sapien
14	441	37.6	219	1	CLD3_MOUSE	Q9z0g9 mus musculu
15	437	37.2	219	1	CLD3_RAT	Q63400 rattus norv
16	437	37.2	220	1	CLD6_HUMAN	P56747 homo sapien
17	434	37.0	217	1	CLD9_MOUSE	Q9z0s7 mus musculu
18	431.5	36.8	211	1	CLD7_MOUSE	Q9z261 mus musculu
19	427.5	36.4	211	1	CLD7_HUMAN	O95471 homo sapien
20	407.5	34.7	211	1	CLD1_HUMAN	O95832 homo sapien
21	401.5	34.2	211	1	CLD1_MOUSE	O88551 mus musculu
22	400	34.1	209	1	CLDZ_BRARE	Q9yh90 brachydanio
23	399.5	34.0	215	1	CLDX_BRARE	Q9yh92 brachydanio
24	391.5	33.3	219	1	CLDJ_HUMAN	P56880 homo sapien
25	387.5	33.0	211	1	CLD1_RAT	P56745 rattus norv
26	386.5	32.9	225	1	CLD8_MOUSE	Q9z260 mus musculu
27	378	32.2	218	1	CLD5_HUMAN	O00501 homo sapien
28	376	32.0	224	1	CLDH_HUMAN	P56750 homo sapien
29	375.5	32.0	191	1	CLD7_RAT	Q9z111 rattus norv
30	372.5	31.7	225	1	CLD8_HUMAN	P56748 homo sapien
31	367	31.3	218	1	CLD5_MOUSE	O54942 mus musculu
32	353	30.1	206	1	CLD5_RAT	Q9jkd6 rattus norv
33	327.5	27.9	228	1	CLDF_HUMAN	P56746 homo sapien
34	291	24.8	228	1	CLDA_HUMAN	P78369 homo sapien
35	290	24.7	231	1	CLDA_MOUSE	Q9z0s6 mus musculu
36	274	23.3	264	1	CLDI_MOUSE	P56857 mus musculu
37	264.5	22.5	261	1	CLDI_HUMAN	P56856 homo sapien
38	223.5	19.0	211	1	CLDD_MOUSE	Q9z0s4 mus musculu
39	192.5	16.4	207	1	CLDB_HUMAN	O75508 homo sapien
40	185.5	15.8	207	1	CLDB_MOUSE	Q60771 mus musculu
41	154.5	13.2	122	1	CLDF_MOUSE	Q9z0s5 mus musculu
42	153.5	13.1	254	1	CLDG_BOVIN	Q9xt98 bos taurus
43	149.5	12.7	305	1	CLDG_HUMAN	Q9y5i7 homo sapien
44	136.5	11.6	244	1	CLDC_HUMAN	P56749 homo sapien
45	101.5	8.6	323	1	CCG2_HUMAN	Q9y698 homo sapien

ALIGNMENTS

```
RESULT   1
CLD2_HUMAN
ID   CLD2_HUMAN     STANDARD;      PRT;   230 AA.
AC   P57739;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Claudin-2.
GN   CLDN2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Epithelium;
RA   Reinecker H.-C., Sakaguchi T., Golden H.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA   Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA   Yu J., Han L.H.;
RT   "Novel human cDNA clone with function of inhibiting cancer cell
RT   growth.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Kidney;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF250558; AAF98151.1; -.
DR   EMBL; AF177340; AAG17984.1; -.
DR   EMBL; BC015252; AAH15252.1; -.
DR   EMBL; BC014424; AAH14424.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
```

```
FT   TRANSMEM    163    183       POTENTIAL.
SQ   SEQUENCE   230 AA;  24548 MW;  52CA642D4A62B70D CRC64;


  Query Match             100.0%;  Score 1174;  DB 1;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 7e-86;
  Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230


RESULT   2
CLD2_CANFA
ID   CLD2_CANFA     STANDARD;      PRT;   230 AA.
AC   Q95KM6;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Claudin-2.
GN   CLDN2.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=21206012; PubMed=11309408;
RA   Furuse M., Furuse K., Sasaki H., Tsukita S.;
RT   "Conversion of zonulae occludentes from tight to leaky strand type by
RT   introducing claudin-2 into Madin-Darby canine kidney I cells.";
RL   J. Cell Biol. 153:263-272(2001).
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF358907; AAK51433.1; -.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    163    183       POTENTIAL.
SQ   SEQUENCE   230 AA;  24502 MW;  91B71C1E5CDC4BE9 CRC64;


  Query Match             95.3%;  Score 1119;  DB 1;  Length 230;
  Best Local Similarity   93.5%;  Pred. No. 1.5e-81;
  Matches  215;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||||||||||||||||||||||||||||||||||||||:||||| ||||||||||||||||||||
Db       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWRTSSYVGTSIVTAVGFSKGLWMECATHSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||||||||:|||||||||||:||||||:||
Db      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIVSVVGMRCTVFCQDSRAKDRLAVV 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||||||:|||||||||||||||||||||||||||||||||||||||||||||:||||
Db     121 GGVFFIIGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLVAGII 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           |||||  | |||:|||:||||||||| |||||||| ||||||||||||
Db     181 LCFSCPLQGNRSDYYDSYQAQPLATRGSPRPGQPPKAKSEFNSYSLTGYV 230


RESULT   3
CLD2_MOUSE
ID   CLD2_MOUSE     STANDARD;      PRT;   230 AA.
AC   O88552;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Claudin-2.
GN   CLDN2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98311639; PubMed=9647647;
RA   Furuse M., Fujita K., Hiiragi T., Fujimoto K., Tsukita S.;
RT   "Claudin-1 and -2: novel integral membrane proteins localizing at
RT   tight junctions with no sequence similarity to occludin.";
RL   J. Cell Biol. 141:1539-1550(1998).
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072128; AAC27079.1; -.
DR   MGD; MGI:1276110; Cldn2.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    163    183       POTENTIAL.
SQ   SEQUENCE   230 AA;  24483 MW;  38A7C074A1E0D5D2 CRC64;


  Query Match             93.4%;  Score 1096;  DB 1;  Length 230;
  Best Local Similarity   91.3%;  Pred. No. 9.9e-80;
  Matches  210;  Conservative   11;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           |||||:||||||||||||| :|||||:|:||||||||||||||||||||||||||||||||
Db       1 MASLGVQLVGYILGLLGLLGTSIAMLLPNWRTSSYVGASIVTAVGFSKGLWMECATHSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           ||||||||||||||||||||||||||||||:||||||||||||||||||:|||||||||
```

```
Db     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAMSSLACIISVVGMRCTVFCQDSRAKDRVAVV 120

Qy    121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
          |||||||||:||||||||||||||||||||||||||||||||||||||:||||:|||:|
Db    121 GGVFFILGGILGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISALFSLVAGVI 180

Qy    181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
          |||||| | ||:|||| ||||||||||||  | || ||||||||||||
Db    181 LCFSCSPQGNRTNYYDGYQAQPLATRSSPRSAQQPKAKSEFNSYSLTGYV 230


RESULT   4
CLDE_HUMAN
ID   CLDE_HUMAN     STANDARD;      PRT;   239 AA.
AC   O95500;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Claudin-14.
GN   CLDN14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Keen T.J., Inglehearn C.F.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT DFNB29 ASP-85.
RC   TISSUE=Liver;
RX   MEDLINE=21097730; PubMed=11163249;
RA   Wilcox E.R., Burton Q.L., Naz S., Riazuddin S., Smith T.N.,
RA   Ploplis B., Belyantseva I., Ben-Yosef T., Liburd N.A., Morell R.J.,
RA   Kachar B., Wu D.K., Griffith A.J., Riazuddin S., Friedman T.B.;
RT   "Mutations in the gene encoding tight junction claudin-14 cause
RT   autosomal recessive deafness DFNB29.";
RL   Cell 104:165-172(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: LIVER, KIDNEY. ALSO FOUND IN EAR.
CC   -!- DISEASE: DEFECTS IN CLDN14 ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNB29).
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ132445; CAA10669.1; -.
DR   EMBL; AF314090; AAG60052.1; -.
DR   EMBL; BC012126; AAH12126.1; -.
DR   EMBL; AP001726; BAA95509.1; -.
DR   MIM; 605608; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane; Disease mutation; Deafness.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    116    136       POTENTIAL.
FT   TRANSMEM    163    183       POTENTIAL.
FT   VARIANT      85     85       V -> D (IN DFNB29).
FT                                /FTId=VAR_010738.
SQ   SEQUENCE   239 AA;  25699 MW;  DD41652F7FD0E09A CRC64;


  Query Match             48.6%;  Score 571;  DB 1;  Length 239;
  Best Local Similarity   46.8%;  Pred. No. 2.4e-38;
  Matches  111;  Conservative   44;  Mismatches   70;  Indels   12;  Gaps    3;

Qy      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
          |||  :||:|::|  ||::|||:  :|| |: :::|| :|:||| : |||||||  ||||
Db      1 MASTAVQLLGFLLSFLGMVGTLITTILPHWRRTAHVGTNILTAVSYLKGLWMECVWHSTG 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
          | || || :|| || |:|||:|:|| |  :| :||  :|:||:||   : : ||   |:
Db     61 IYQCQIYRSLLALPQDLQAARALMVISCLLSGIACACAVIGMKCTRCAKGTPAKTTFAIL 120

Qy    121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
          ||  ||| |||  : |:|  : ::::||:||:|  ||||||:||||| |||  ||| | :
Db    121 GGTLFILAGLLCMVAVSWTTNDVVQNFYNPLLPSGMKFEIGQALYLGFISSSLSLIGGTL 180

Qy    181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPG---QPPKVKSEFNSYSLT-----GY 229
          || ||  :      |  ||| | || ::       |||    :  : |:|     ||
Db    181 LCLSCQDEAP----YRPYQAPPRATTTTANTAPAYQPPAAYKDNRAPSVTSATHSGY 233


RESULT   5
CLDE_MOUSE
ID   CLDE_MOUSE     STANDARD;      PRT;   239 AA.
AC   Q9Z0S3; Q9D284;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Claudin-14.
GN   CLDN14.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Morita K., Furuse M., Tsukita S.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Brain;
RX   MEDLINE=21097730; PubMed=11163249;
RA   Wilcox E.R., Burton Q.L., Naz S., Riazuddin S., Smith T.N.,
RA   Ploplis B., Belyantseva I., Ben-Yosef T., Liburd N.A., Morell R.J.,
RA   Kachar B., Wu D.K., Griffith A.J., Riazuddin S., Friedman T.B.;
RT   "Mutations in the gene encoding tight junction claudin-14 cause
RT   autosomal recessive deafness DFNB29.";
RL   Cell 104:165-172(2001).
RN   [3]
RP   REVISIONS TO 115; 129; 166 AND 187.
```

```
RA   Wilcox E.R., Burton Q.L., Naz S., Riazuddin S., Smith T.N.,
RA   Ploplis B., Belyantseva I., Ben-Yosef T., Liburd N.A., Morell R.J.,
RA   Kachar B., Wu D.K., Griffith A.J., Riazuddin S., Friedman T.B.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Colon;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: AT POSTNATAL DAY 4, EXPRESSION IS APICALLY
CC       LOCATED IN THE INNER AND OUTER HAIR CELL REGION OF THE ENTIRE
CC       ORGAN OF CORTI. BY POSTNATAL DAY 8, EXPRESSION IS HIGHEST IN THE
CC       SUPPORTING CELLS OF THE ORGAN OF CORTI AND PRESENT IN ALL OTHER
CC       SENSORY EPITHELIA OF THE INNER EAR VESTIBULAR ORGANS. ALSO
CC       EXPRESSED IN LIVER AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF124429; AAD17323.1; -.
DR   EMBL; AF314089; AAG60051.2; -.
DR   EMBL; AK020255; BAB32041.1; -.
DR   MGD; MGI:1860425; Cldn14.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    116    136       POTENTIAL.
FT   TRANSMEM    163    183       POTENTIAL.
FT   CONFLICT    115    115       T -> N (IN REF. 1).
FT   CONFLICT    119    119       V -> M (IN REF. 4).
FT   CONFLICT    129    129       G -> A (IN REF. 1).
FT   CONFLICT    166    166       L -> M (IN REF. 1).
FT   CONFLICT    187    187       D -> E (IN REF. 1).
SQ   SEQUENCE   239 AA;  25614 MW;  24DE6AEADA56BB61 CRC64;


  Query Match             47.8%;  Score 561.5;  DB 1;  Length 239;
  Best Local Similarity   46.9%;  Pred. No. 1.4e-37;
  Matches  112;  Conservative   43;  Mismatches   75;  Indels    9;  Gaps    3;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
```

```
Db       1 MASTAVQLLGFLLSFLGMVGTLITTILPHWRRTAHVGTNILTAVSYLKGLWMECVWHSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           | || || :|| || |:|||:|:|| |  :| :||  :||||:||    : : ||    ||
Db      61 IYQCQIYRSLLALPRDLQAARALMVISCLLSGMACACAVVGMKCTRCAKGTPAKTTFAVL 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||  |:| |||  : |:|  : ::::||:||:|  ||||||:||||| |||  ||| | :
Db     121 GGALFLLAGLLCMVAVSWTTNDVVQNFYNPLLPSGMKFEIGQALYLGFISSSLSLIGGTL 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSP--RP------GQPPKVKSEFNS-YSLTGYV 230
           || ||  :    |   :|   | ::|  ||       : | | |  :| | |  ||
Db     181 LCLSCQDEAPYRPYPPQSRAGATTTATAPAYRPPAAYKDNRAPSVTSAAHSGYRLNDYV 239


RESULT   6
CLDY_BRARE
ID   CLDY_BRARE     STANDARD;      PRT;   208 AA.
AC   Q9YH91;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Claudin-like protein ZF-A89.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Keen T.J., Inglehearn C.F.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ011789; CAA09777.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
SQ   SEQUENCE   208 AA;  22205 MW;  6A4BB5EBF3CCAB81 CRC64;


  Query Match             39.1%;  Score 459.5;  DB 1;  Length 208;
  Best Local Similarity   40.7%;  Pred. No. 1.3e-29;
  Matches   85;  Conservative   49;  Mismatches   70;  Indels    5;  Gaps    1;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           |||:||||:  :| ::| || :|   || || ::::| :|||| | :|||| |   |||
Db       1 MASVGLQLLATVLAIIGWLGEIVICALPMWKVTAFIGNNIVTAQIFWEGLWMNCVQQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             || :| ::| || |:|||:|::| |  :: :   ::: | :||   ::  || :| ||
Db      61 QMQCKVYDSMLALPQDLQAARALVVISIIVTFMGVFLTIAGGKCTNCIEDQDAKAKVVVA 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||||::||:|  ||| |: : :::|||:| : |: | |:| :|::|  :|   |: | :
```

```
Db    121 AGVFFLVGGILCLIPVCWSANSVIKDFYNPTLSDAQKRELGASLFIGWCASGLLLLGGAL 180

Qy    181 LCFSCSSQRNRSNYYDAYQAQPLATRSSP 209
          || |      |     ||  :  | ||:|
Db    181 LCCQCPKNEGR-----AYSVKYSAPRSAP 204


RESULT   7
CLD3_CANFA
ID   CLD3_CANFA     STANDARD;      PRT;   218 AA.
AC   Q95KM5;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Claudin-3.
GN   CLDN3.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=21206012; PubMed=11309408;
RA   Furuse M., Furuse K., Sasaki H., Tsukita S.;
RT   "Conversion of zonulae occludentes from tight to leaky strand type by
RT   introducing claudin-2 into Madin-Darby canine kidney I cells.";
RL   J. Cell Biol. 153:263-272(2001).
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF358908; AAK51434.1; -.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    116    136       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
SQ   SEQUENCE   218 AA;  23148 MW;  907104E36F50DA70 CRC64;


  Query Match             38.4%;  Score 450.5;  DB 1;  Length 218;
  Best Local Similarity   39.6%;  Pred. No. 6.9e-29;
  Matches   84;  Conservative   47;  Mismatches   74;  Indels    7;  Gaps    1;

Qy      3 SLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGIT 62
          |:||:: |  | :|| | |:|   || |: ::::|:||:||    :|||| |   |||
Db      2 SMGLEIAGTSLAVLGWLSTIVCCALPMWRVTAFIGSSIITAQITWEGLWMNCVVQSTGQM 61

Qy     63 QCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGG 122
          || :| :|| || |:|||:|::| |  :::   ::::|| :||   |:  || :: :  |
Db     62 QCKVYDSLLALPQDLQAARALIVVSILLAAFGLLVALVGAQCTNCVQDDTAKAKITIVAG 121

Qy    123 VFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILC 182
          | |:|  ||  :||:|: : |:||||:|||||: | |:|  ||:|  ::   |: | :||
Db    122 VLFLLAALLTLVPVSWSANTIIRDFYNPLVPDAQKREMGAGLYVGWAAAALQLLGGALLC 181

Qy    183 FSCSSQRNRSNYYDAYQAQPLATRSSPRPGQP 214
           ||  :       |   |      |:||   |
Db    182 CSCPPR-------DKKYAPTKIVYSAPRSAGP 206


RESULT   8
CLD3_HUMAN
ID   CLD3_HUMAN     STANDARD;      PRT;   220 AA.
AC   O15551;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Claudin-3 (Clostridium perfringens enterotoxin receptor 2) (CPE-
DE   receptor 2) (CPE-R 2) (Ventral prostate.1 protein homolog) (HRVP1).
GN   CLDN3 OR CPETR2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98110580; PubMed=9441748;
RA   Peacock R.E., Keen T.J., Inglehearn C.F.;
RT   "Analysis of a human gene homologous to rat ventral prostate.1
RT   protein.";
RL   Genomics 46:443-449(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97476271; PubMed=9334247;
RA   Katahira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RA   Sugimoto N.;
RT   "Clostridium perfringens enterotoxin utilizes two structurally related
RT   membrane proteins as functional receptors in vivo.";
RL   J. Biol. Chem. 272:26652-26658(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF007189; AAC78277.1; -.
DR   EMBL; AB000714; BAA22986.1; -.
DR   EMBL; BC016056; AAH16056.1; -.
DR   MIM; 602910; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    116    136       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
SQ   SEQUENCE   220 AA;  23318 MW;  1CB26EFFF1563C56 CRC64;


  Query Match             38.3%;  Score 450;  DB 1;  Length 220;
  Best Local Similarity   39.5%;  Pred. No. 7.7e-29;
  Matches   83;  Conservative   51;  Mismatches   72;  Indels    4;  Gaps    2;

Qy      3 SLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGIT 62
          |:||:: |  | :|| |||:|   || |: |:::|::|:|:    :|||| |   |||
Db      2 SMGLEITGTALAVLGWLGTIVCCALPMWRVSAFIGSNIITSQNIWEGLWMNCVVQSTGQM 61

Qy     63 QCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGG 122
          || :| :|| || |:|||:|::| :  :::   ::::|| :||   |:  || :: :  |
Db     62 QCKVYDSLLALPQDLQAARALIVVAILLAAFGLLVALVGAQCTNCVQDDTAKAKITIVAG 121
```

```
Qy    123 VFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILC 182
          | |:|  ||  :||:|: : |:||||:|:||:: | |:|  ||:|  ::   |: | :||
Db    122 VLFLLAALLTLVPVSWSANTIIRDFYNPVVPEAQKREMGAGLYVGWAAAALQLLGGALLC 181

Qy    183 FSCSSQRNRSNYYDAYQAQPLATRSSPRPG 212
            ||     |   | | :    | ||:  ||
Db    182 CSCPP---REKKYTATKVVYSAPRST-GPG 207


RESULT  9
CLD4_MOUSE
ID   CLD4_MOUSE     STANDARD;      PRT;   210 AA.
AC   O35054;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Claudin-4 (Clostridium perfringens enterotoxin receptor) (CPE-
DE   receptor) (CPE-R).
GN   CLDN4 OR CPETR1 OR CPER.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97476271; PubMed=9334247;
RA   Katahira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RA   Sugimoto N.;
RT   "Clostridium perfringens enterotoxin utilizes two structurally related
RT   membrane proteins as functional receptors in vivo.";
RL   J. Biol. Chem. 272:26652-26658(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99110921; PubMed=9892664;
RA   Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT   "Claudin multigene family encoding four-transmembrane domain protein
RT   components of tight junction strands.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB000713; BAA22985.1; -.
DR   EMBL; AF087822; AAD09757.1; -.
DR   MGD; MGI:1313314; Cldn4.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    161    181       POTENTIAL.
SQ   SEQUENCE   210 AA;  22338 MW;  3B6D571EC71D6564 CRC64;


  Query Match             38.0%;  Score 446;  DB 1;  Length 210;
  Best Local Similarity   40.2%;  Pred. No. 1.5e-28;
  Matches   84;  Conservative   46;  Mismatches   75;  Indels    4;  Gaps    1;

Qy      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
            |||:|||::|  | :|| || :::  || |: ::::|::|||||    :|||| |   |||
Db      1 MASMGLQVLGISLAVLGWLGIILSCALPMWRVTAFIGSNIVTAQTSWEGLWMNCVVQSTG 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
            || :| ::| || |:|||:|:|| |  : :|  ::|||| :||   ::   | :: :
Db     61 QMQCKMYDSMLALPQDLQAARALMVISIIVGALGMLLSVVGGKCTNCMEDETVKAKIMIT 120

Qy    121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           |  ||:  :|  :||:|  | ::||||:|:|    | |:| :||:|  :|   |: | :
Db    121 AGAVFIVASMLIMVPVSWTAHNVIRDFYNPMVASGQKREMGASLYVGWAASGLLLLGGGL 180

Qy    181 LCFSCSSQRNRSNYYDAYQAQPLATRSSP 209
          || ||  : |   |   |     | || |
Db    181 LCCSCPPRSNDKPYSAKYS----AARSVP 205


RESULT  10
CLD4_CERAE
ID   CLD4_CERAE     STANDARD;      PRT;   209 AA.
AC   O19005;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Claudin-4 (Clostridium perfringens enterotoxin receptor) (CPE-
DE   receptor) (CPE-R).
GN   CLDN4 OR CPETR1 OR CPER.
OS   Cercopithecus aethiops (Green monkey) (Grivet).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=9534;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97242441; PubMed=9087440;
RA   Katahira J., Inoue N., Horiguchi Y., Matsuda M., Sugimoto N.;
RT   "Molecular cloning and functional characterization of the receptor for
RT   Clostridium perfringens enterotoxin.";
RL   J. Cell Biol. 136:1239-1247(1997).
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D88492; BAA22781.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      10     30       POTENTIAL.
FT   TRANSMEM      82    102       POTENTIAL.
FT   TRANSMEM     118    138       POTENTIAL.
FT   TRANSMEM     161    181       POTENTIAL.
SQ   SEQUENCE   209 AA;  22029 MW;  474DB3099F95289E CRC64;


  Query Match             37.8%;  Score 443.5;  DB 1;  Length 209;
  Best Local Similarity   39.9%;  Pred. No. 2.4e-28;
  Matches   83;  Conservative   49;  Mismatches   71;  Indels    5;  Gaps    1;

Qy      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
            |||:|||: |  | :|| |  ::   || |: ::::|::|||:    :|||| |   |||
Db      1 MASMGLQVTGIALAVLGWLAVMLCCALPMWRVTAFIGSNIVTSQTIWEGLWMNCVVQSTG 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
            || :| :|| || |:|||:|::: |  :::|  ::|||| :||   ::  || :  :
Db     61 QMQCKVYDSLLALPQDLQAARALVIISIIVAALGVLLSVVGGKCTNCLEDESAKAKTMIV 120
```

```
Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           || |:| |||  :||:|  | |::|||:|||    | |:| :||:|  :|   |: | :
Db     121 AGVVFLLAGLLVIVPVSWTAHNIIQDFYNPLVASGQKREMGASLYVGWAASGLLLLGGGL 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSS 208
           || :|  : ::      | |:  | ||:
Db     181 LCCNCPPRTDK-----PYSAKYSAARSA 203


RESULT  11
CLD6_MOUSE
ID   CLD6_MOUSE     STANDARD;      PRT;   219 AA.
AC   Q9Z262;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Claudin-6.
GN   CLDN6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99110921; PubMed=9892664;
RA   Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT   "Claudin multigene family encoding four-transmembrane domain protein
RT   components of tight junction strands.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF087824; AAD09759.1; -.
DR   EMBL; BC005718; AAH05718.1; -.
DR   MGD; MGI:1859284; Cldn6.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    164    184       POTENTIAL.
SQ   SEQUENCE   219 AA;  23387 MW;  00CF687FAB687E1F CRC64;


  Query Match             37.7%;  Score 443;  DB 1;  Length 219;
  Best Local Similarity   43.9%;  Pred. No. 2.7e-28;
  Matches   82;  Conservative   39;  Mismatches   66;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
             ||| |||::| :| ||| :  ||:  || || ::::| ||| |    :|||| |   |||
Db       1 MASTGLQILGIVLTLLGWVNALVSCALPMWKVTAFIGNSIVVAQMVWEGLWMSCVVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
              || :| :|| || |:|||:|: | :  |  |  :: : | :||   ::  :| |: :
Db      61 QMQCKVYDSLLALPQDLQAARALCVVTLLIVLLGLLVYLAGAKCTTCVEDRNSKSRLVLI 120
```

```
Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           |: |:: |:|  ||| |  | |::|||:||| |: | |:| :||||  :|   |: | :
Db     121 SGIIFVISGVLTLIPVCWTAHSIIQDFYNPLVADAQKRELGASLYLGWAASGLLLLGGGL 180

Qy     181 LCFSCSS 187
           || :|||
Db     181 LCCACSS 187


RESULT  12
CLD4_HUMAN
ID   CLD4_HUMAN     STANDARD;      PRT;   209 AA.
AC   O14493;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Claudin-4 (Clostridium perfringens enterotoxin receptor) (CPE-
DE   receptor) (CPE-R).
GN   CLDN4 OR CPETR1 OR CPER.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=97476271; PubMed=9334247;
RA   Katahira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RA   Sugimoto N.;
RT   "Clostridium perfringens enterotoxin utilizes two structurally related
RT   membrane proteins as functional receptors in vivo.";
RL   J. Biol. Chem. 272:26652-26658(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB000712; BAA22984.1; -.
DR   EMBL; BC000671; AAH00671.1; -.
DR   MIM; 602909; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    118    138       POTENTIAL.
FT   TRANSMEM    161    181       POTENTIAL.
SQ   SEQUENCE   209 AA;  22077 MW;  0659A93AA5F0E4C5 CRC64;


  Query Match             37.7%;  Score 442.5;  DB 1;  Length 209;
  Best Local Similarity   39.4%;  Pred. No. 2.8e-28;
  Matches   82;  Conservative   51;  Mismatches   70;  Indels    5;  Gaps    1;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
             |||:|||::|  | :|| |  ::   || |: ::::|::|||:    :|||| |   |||
Db       1 MASMGLQVMGIALAVLGWLAVMLCCALPMWRVTAFIGSNIVTSQTIWEGLWMNCVVQSTG 60
```

```
Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           || :| :|| || |:|||:|::: |  :::|  ::|||| :||   ::  || :  :
Db      61 QMQCKVYDSLLALPQDLQAARALVIISIIVAALGVLLSVVGGKCTNCLEDESAKAKTMIV 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           || |:| ||:  :||:|  | |::|||:|||    | |:| :||:|  :|   |: | :
Db     121 AGVVFLLAGLMVIVPVSWTAHNIIQDFYNPLVASGQKREMGASLYVGWAASGLLLLGGGL 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSS 208
           || :|  : ::      | |:  | ||:
Db     181 LCCNCPPRTDK-----PYSAKYSAARSA 203


RESULT  13
CLD9_HUMAN
ID   CLD9_HUMAN     STANDARD;      PRT;   217 AA.
AC   O95484;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Claudin-9.
GN   CLDN9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Keen T.J., Inglehearn C.F.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ130941; CAA10254.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
SQ   SEQUENCE   217 AA;  22848 MW;  0E49C5B5CB3AAC9C CRC64;


  Query Match             37.6%;  Score 441;  DB 1;  Length 217;
  Best Local Similarity   43.2%;  Pred. No. 3.9e-28;
  Matches   80;  Conservative   42;  Mismatches   63;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
             ||| ||:|:|  | :|| |||||:  || || ::::| ||| |    :|||| |   |||
Db       1 MASTGLELLGMTLAVLGWLGTLVSCALPLWKVTAFIGNSIVVAQVVWEGLWMSCVVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
           || :| :|| || |:|||:|: | :  :: |  :::: | :||   ::  || |: :
Db      61 QMQCKVYDSLLALPQDLQAARALCVIALLLALLGLLVAITGAQCTTCVEDEGAKARIVLT 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
             ||  :| |:|  ||| |  | |::|||:||| :::| |:| :||||  ::   :: | :
Db     121 AGVILLLAGILVLIPVCWTAHAIIQDFYNPLVAEALKRELGASLYLGWAAAALLMLGGGL 180

Qy     181 LCFSC 185
           || :|
Db     181 LCCTC 185


RESULT  14
CLD3_MOUSE
ID   CLD3_MOUSE     STANDARD;      PRT;   219 AA.
AC   Q9Z0G9;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Claudin-3 (Clostridium perfringens enterotoxin receptor 2) (CPE-
DE   receptor 2) (CPE-R 2).
GN   CLDN3 OR CPETR2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99097345; PubMed=9878248;
RA   Paperna T., Peoples R., Wang Y.K., Kaplan P., Francke U.;
RT   "Genes for the CPE receptor (CPETR1) and the human homolog of RVP1
RT   (CPETR2) are localized within the Williams-Beuren syndrome deletion.";
RL   Genomics 54:453-459(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99110921; PubMed=9892664;
RA   Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT   "Claudin multigene family encoding four-transmembrane domain protein
RT   components of tight junction strands.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF095905; AAD14608.1; -.
DR   EMBL; AF087821; AAD09756.1; -.
DR   MGD; MGI:1329044; Cldn3.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    116    136       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
SQ   SEQUENCE   219 AA;  23284 MW;  62F67810D9B9BD37 CRC64;


  Query Match             37.6%;  Score 441;  DB 1;  Length 219;
  Best Local Similarity   40.1%;  Pred. No. 3.9e-28;
  Matches   85;  Conservative   47;  Mismatches   72;  Indels    8;  Gaps    2;

Qy       3 SLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGIT 62
             |:||:: |  | :|| | |:|   || |: |:::|:||:||    :|||| |   |||
Db       2 SMGLEITGTSLAVLGWLCTIVCCALPMWRVSAFIGSSIITAQITWEGLWMNCVVQSTGQM 61

Qy      63 QCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGG 122
           || :| :|| || |:|||:|::| |  :::   ::::|| :||   |:  || :: :  |
Db      62 QCKMYDSLLALPQDLQAARALIVVSILLAAFGLLVALVGAQCTNCVQDETAKAKITIVAG 121
```

```
Qy    123 VFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILC 182
          | |:|  ||  :||:|: : |:||||:||||:: | |:|  ||:|  ::   |: | :||
Db    122 VLFLLAALLTLVPVSWSANTIIRDFYNPLVPEAQKREMGAGLYVGWAAAALQLLGGALLC 181

Qy    183 FSCSSQRNRSNYYDAYQAQPLATRSSPRPGQP 214
           ||  :      | | |      |:||   |
Db    182 CSCPPR-------DKY-APTKILYSAPRSTGP 205


RESULT 15
CLD3_RAT
ID   CLD3_RAT       STANDARD;      PRT;   219 AA.
AC   Q63400;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Claudin-3 (Ventral prostate.1 protein) (RVP1).
GN   CLDN3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92130987; PubMed=1723140;
RA   Briehl M.M., Miesfeld R.L.;
RT   "Isolation and characterization of transcripts induced by androgen
RT   withdrawal and apoptotic cell death in the rat ventral prostate.";
RL   Mol. Endocrinol. 5:1381-1388(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Keen T.J., Inglehearn C.F.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M74067; AAA41760.1; -.
DR   EMBL; AJ011656; CAA09727.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; 1.
KW   Tight junction; Transmembrane.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    116    136       POTENTIAL.
FT   TRANSMEM    162    182       POTENTIAL.
FT   CONFLICT      4      4       G -> S (IN REF. 1).
FT   CONFLICT     55     55       MISSING (IN REF. 1).
FT   CONFLICT    217    219       DYV -> TTSERPGARTPHHHHYQPSMYPTRPACSLASETT
FT                                PPSRRLQTPRSLLARLEEDRQPGVPFSPVAT (IN REF.
FT                                1).
SQ   SEQUENCE   219 AA;  23314 MW;  820CC6BFC20D122D CRC64;


  Query Match             37.2%;  Score 437;  DB 1;  Length 219;
  Best Local Similarity   39.6%;  Pred. No. 8.1e-28;
  Matches   84;  Conservative   47;  Mismatches   73;  Indels    8;  Gaps    2;

Qy      3 SLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGIT 62
            |:||:: |  | :|| | |:|   || |: |:::|:||:||    :|||| |   |||
Db      2 SMGLEITGTSLAVLGWLCTIVCCALPMWRVSAFIGSSIITAQITWEGLWMNCVVQSTGQM 61

Qy     63 QCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGG 122
          || :| :|| || |:|||:|::| |  :::   ::::|| :||   |:  || :: :  |
Db     62 QCKMYDSLLALPQDLQAARALIVVSILLAAFGLLVALVGAQCTNCVQDETAKAKITIVAG 121

Qy    123 VFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILC 182
          | |:|  :|  :||:|: : |:||||:||||:: | |:|  ||:|  ::   |: | :||
Db    122 VLFLLAAVLTLVPVSWSANTIIRDFYNPLVPEAQKREMGTGLYVGWAAAALQLLGGALLC 181

Qy    183 FSCSSQRNRSNYYDAYQAQPLATRSSPRPGQP 214
           ||     |  |     |      |:||   |
Db    182 CSCPP---REKY-----APTKILYSAPRSTGP 205


Search completed: June  2, 2002, 20:17:10
Job time: 478 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 2, 2002, 20:00:17 ; Search time 108.83 Seconds
(without alignments)
365.605 Million cell updates/sec

Title: US-09-663-600A-186
Perfect score: 1174
Sequence: 1 MASLGLQLVGYILGLLGLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1119	95.3	230	6	Q95KM6	Q95km6 canis famil
2	456	38.8	218	13	Q90XR2	Q90xr2 brachydanio
3	450.5	38.4	218	6	Q95KM5	Q95km5 canis famil
4	446	38.0	215	13	Q90XR8	Q90xr8 brachydanio
5	444.5	37.9	209	13	Q90XR9	Q90xr9 brachydanio
6	438	37.3	210	13	Q90XQ8	Q90xq8 brachydanio
7	432.5	36.8	214	13	Q90XQ9	Q90xq9 xenopus lae
8	430	36.6	214	13	Q9DE12	Q9de12 xenopus lae
9	428.5	36.5	214	13	Q98SR2	Q98sr2 gallus gall
10	421	35.9	214	13	Q90XR0	Q90xr0 brachydanio
11	417.5	35.6	210	13	Q90XR4	Q90xr4 brachydanio
12	413.5	35.2	206	13	Q90XS1	Q90xs1 brachydanio
13	396	33.7	211	13	Q90XR6	Q90xr6 brachydanio
14	393.5	33.5	193	11	Q9ET38	Q9et38 mus musculu
15	393	33.5	216	13	Q98SR1	Q98sr1 gallus gall
16	375.5	32.0	200	11	Q91X40	Q91x40 mus musculu
17	328.5	28.0	237	13	Q90XS0	Q90xs0 brachydanio
18	276.5	23.6	224	5	Q9N9W2	Q9n9w2 halocynthia
19	276.5	23.6	224	5	Q9N9W1	Q9n9w1 halocynthia
20	276	23.5	227	11	Q9D7Z0	Q9d7z0 mus musculu
21	270	23.0	208	11	Q91ZZ1	Q91zz1 mus musculu
22	270	23.0	227	11	Q9D8A6	Q9d8a6 mus musculu
23	269.5	23.0	261	4	Q96PH4	Q96ph4 homo sapien
24	264	22.5	264	11	Q91ZZ0	Q91zz0 mus musculu
25	260.5	22.2	226	4	Q96N78	Q96n78 homo sapien
26	260	22.1	208	11	Q91ZY9	Q91zy9 mus musculu
27	256	21.8	219	13	Q90XR7	Q90xr7 brachydanio
28	247	21.0	229	11	Q9CX57	Q9cx57 mus musculu
29	218.5	18.6	220	11	Q9D7U6	Q9d7u6 mus musculu
30	191	16.3	256	13	Q90XR5	Q90xr5 brachydanio
31	182.5	15.5	207	11	Q9DB65	Q9db65 mus musculu
32	182	15.5	296	11	Q9D7D7	Q9d7d7 mus musculu
33	177.5	15.1	207	11	Q99P82	Q99p82 rattus norv
34	153.5	13.1	235	6	Q9TUF7	Q9tuf7 bos taurus
35	153	13.0	268	4	Q96B33	Q96b33 homo sapien
36	151	12.9	235	11	Q91Y55	Q91y55 rattus norv
37	150	12.8	341	13	Q90XR3	Q90xr3 brachydanio
38	150	12.8	341	13	Q90XR1	Q90xr1 brachydanio
39	148	12.6	235	11	Q925N4	Q925n4 mus musculu
40	137	11.7	210	11	Q9D5B8	Q9d5b8 mus musculu
41	133	11.3	209	11	Q9D9N2	Q9d9n2 mus musculu
42	130	11.1	228	11	Q9ET43	Q9et43 mus musculu
43	116	9.9	128	4	Q96FX9	Q96fx9 homo sapien
44	114.5	9.8	200	4	Q9Y693	Q9y693 homo sapien
45	109	9.3	234	11	Q9CXT3	Q9cxt3 mus musculu

ALIGNMENTS

RESULT 1
Q95KM6
ID Q95KM6 PRELIMINARY; PRT; 230 AA.
AC Q95KM6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTEGRAL MEMBRANE PROTEIN CLAUDIN-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=21206012; PubMed=11309408;
RA Furuse M., Furuse K., Sasaki H., Tsukita S.;
RT "Conversion of Zonulae Occludentes from Tight to Leaky Strand Type by
RT Introducing Claudin-2 into Madin-Darby Canine Kidney I Cells.";
RL J. Cell Biol. 153:263-272(2001).
DR EMBL; AF358907; AAK51433.1; -.
SQ SEQUENCE 230 AA; 24502 MW; 91B71C1E5CDC4BE9 CRC64;

Query Match 95.3%; Score 1119; DB 6; Length 230;
Best Local Similarity 93.5%; Pred. No. 5.9e-90;
Matches 215; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

```
Qy      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
          ||||||||||||||||||||||||||||:|||||| ||||||||||||||||||||||
Db      1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWRTSSYVGTSIVTAVGFSKGLWMECATHSTG 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
          ||||||||||||||||||||||||||||||||||:||||||||||:||||||:|||||:||
Db     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIVSVVGMRCTVFCQDSRAKDRLAVV 120

Qy    121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
          ||||||:||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    121 GGVFFIIGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLVAGII 180
```

```
Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           |||||  | |||:|||:|||||||||| |||||||| ||||||||||||
Db     181 LCFSCPLQGNRSDYYDSYQAQPLATRGSPRPGQPPKAKSEFNSYSLTGYV 230


RESULT   2
Q90XR2
ID   Q90XR2      PRELIMINARY;      PRT;   218 AA.
AC   Q90XR2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN C.
GN   CLDNC.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359432; AAL01839.1; -.
SQ   SEQUENCE   218 AA;  23425 MW;  82CA23D13306110F CRC64;


  Query Match             38.8%;  Score 456;  DB 13;  Length 218;
  Best Local Similarity   38.4%;  Pred. No. 3.4e-32;
  Matches   88;  Conservative   42;  Mismatches   71;  Indels   28;  Gaps    1;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
             ||| ||:|||  | :|| :  :|   || |: ::::| :|||| : :|:|| |   |||
Db       1 MASFGLELVGVTLSVLGWILNIVCCALPMWRVTAFIGTNIVTAQVYWEGIWMSCVVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             || :| ::| ||||:|||:|::| :  :  ||  :::|| :||   :|  || || ::
Db      61 QMQCKVYDSMLALPADLQAARALVVVAIIVGVLALFVAIVGAKCTNCIEEEAAKARVMIS 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
             |  ||   :|  ||| |: | :: :|||||||:: | ||| :|||  :|   |: | |
Db     121 SGAAFITASVLQLIPVCWSAHTVILEFYSPLVPEAQKMEIGASLYLGWAASAMLLVGGSI 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGY 229
           || ||                            ||| :: :   |   |
Db     181 LCCSC----------------------------PPKDETRYPPQSRIAY 201


RESULT   3
Q95KM5
ID   Q95KM5      PRELIMINARY;      PRT;   218 AA.
AC   Q95KM5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   INTEGRAL MEMBRANE PROTEIN CLAUDIN-3.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=21206012; PubMed=11309408;
RA   Furuse M., Furuse K., Sasaki H., Tsukita S.;
RT   "Conversion of Zonulae Occludentes from Tight to Leaky Strand Type by
RT   Introducing Claudin-2 into Madin-Darby Canine Kidney I Cells.";
RL   J. Cell Biol. 153:263-272(2001).
DR   EMBL; AF358908; AAK51434.1; -.
SQ   SEQUENCE   218 AA;  23148 MW;  907104E36F50DA70 CRC64;


  Query Match             38.4%;  Score 450.5;  DB 6;  Length 218;
  Best Local Similarity   39.6%;  Pred. No. 1e-31;
  Matches   84;  Conservative   47;  Mismatches   74;  Indels    7;  Gaps    1;

Qy       3 SLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGIT 62
             |:||:: |  | :|| | |:|   || |: ::::|:||:||    :|||| |   |||
Db       2 SMGLEIAGTSLAVLGWLSTIVCCALPMWRVTAFIGSSIITAQITWEGLWMNCVVQSTGQM 61

Qy      63 QCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGG 122
             || :| :|| || |:|||:|::| |  :::   ::::|| :||   |:  || :: :  |
Db      62 QCKVYDSLLALPQDLQAARALIVVSILLAAFGLLVALVGAQCTNCVQDDTAKAKITIVAG 121

Qy     123 VFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILC 182
           | |:|  ||  :||:|: : |:||||:|||||: | |:|  ||:|  ::   |: | :||
Db     122 VLFLLAALLTLVPVSWSANTIIRDFYNPLVPDAQKREMGAGLYVGWAAAALQLLGGALLC 181

Qy     183 FSCSSQRNRSNYYDAYQAQPLATRSSPRPGQP 214
             ||  :       |   |       |:||   |
Db     182 CSCPPR-------DKKYAPTKIVYSAPRSAGP 206


RESULT   4
Q90XR8
ID   Q90XR8      PRELIMINARY;      PRT;   215 AA.
AC   Q90XR8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN B.
GN   CLDNB.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359426; AAL01833.1; -.
SQ   SEQUENCE   215 AA;  22656 MW;  F3D944F527751145 CRC64;


  Query Match             38.0%;  Score 446;  DB 13;  Length 215;
  Best Local Similarity   39.7%;  Pred. No. 2.5e-31;
  Matches   83;  Conservative   45;  Mismatches   81;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
             ||| |||::|  | : | :| :|   || || ::::||:|||:    :|:|| |   |||
Db       1 MASTGLQMLGIALAIFGWIGVIVLCALPMWKVTAFIGANIVTSQTSWEGIWMSCVVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             || :| ::| | :|||||:|: | |  |  :  ::|: | :||   :|  :| :| :  :
Db      61 QMQCKVYDSMLALSSDIQAARALTVISIVIGVMGIMLSMAGGKCTNCIEEESSKAKVGIT 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
             || ||: |:|  :|| |  : |::|||:|||  : | ||| :||:|  :|   :| | :
Db     121 AGVIFIISGVLCLVPVCWTANAIIQDFYNPLVVQAQKREIGASLYIGWGASALLIIGGSL 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSP 209
           || |  : :   |   | | | :  |:|
Db     181 LCCHCPEKSDSGKYTAKYNATPRSEASAP 209


RESULT   5
Q90XR9
ID   Q90XR9      PRELIMINARY;      PRT;   209 AA.
AC   Q90XR9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN E.
GN   CLDNE.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359425; AAL01832.1; -.
SQ   SEQUENCE   209 AA;  22593 MW;  A86F48D852E16CB2 CRC64;

  Query Match             37.9%;  Score 444.5;  DB 13;  Length 209;
  Best Local Similarity   40.4%;  Pred. No. 3.3e-31;
  Matches   84;  Conservative   49;  Mismatches   72;  Indels    3;  Gaps    1;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           | |:  :::|  | ::| || ::   || || ::::||:|||    :|||| |   |||
Db       1 MVSMCREILGMCLAIIGFLGAIIICALPMWKVTAFIGANIVTAQTIWEGLWMNCVMQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             || || :|| || |:|||:|::| :  :|  | |: : | :|| | :   || :|::|
Db      61 QMQCKIYDSLLALPQDLQAARALVVIAIIVSFFALILGIAGGKCTNFVEREDAKAKVSIA 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
            || ||: |:|  :|| |: : |:||||:||: |: : |:| :||:|   :   :| | |
Db     121 SGVIFIIAGVLVLVPVCWSANTIIRDFYNPLLTDAQRREMGASLYIGWGVAALLIIGGGI 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSS 208
           || ||  :  :   |:   :|| :| :|
Db     181 LCSSCPPKDEK---YNMKYSQPRSTATS 205


RESULT   6
Q90XQ8
ID   Q90XQ8      PRELIMINARY;      PRT;   210 AA.
AC   Q90XQ8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN 19.
GN   CLDN19.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359436; AAL01843.1; -.
SQ   SEQUENCE   210 AA;  22253 MW;  FEFFB51CE7A98D47 CRC64;

  Query Match             37.3%;  Score 438;  DB 13;  Length 210;
  Best Local Similarity   41.3%;  Pred. No. 1.2e-30;
  Matches   88;  Conservative   37;  Mismatches   80;  Indels    8;  Gaps    2;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||  |||::|| || ||||| : :  :  || ||| | :|:||    :|||  | : |||
Db       1 MAHAGLQMLGYCLGFLGLLGLIASTAMAEWKMSSYAGDNIITAQAQYEGLWQSCVSQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQES-RAKDRVAV 119
             ||  | :|| || :|| |: :|:|   :  |: ::| |||:||    |: : | :||:
Db      61 QLQCKKYDSLLKLPGEIQGARGLMLTGIFLCGLSTLVSFVGMKCTTCLSEAPQVKSKVAL 120

Qy     120 AGGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGI 179
           |||| || |||   |  :|   | : |: |  | : ::| |:|||:|  ||  |:| |
Db     121 AGGVLFITGGLFALIATSWYGEKIRQKFFDPFTPTNARYEFGKALYVGWGSSALSIIGGS 180

Qy     180 ILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPG 212
           :||  | |: :   |          |:: |||
Db     181 LLCCICGSEASEKPSYP-------PARAAGRPG 206


RESULT   7
Q90XQ9
ID   Q90XQ9      PRELIMINARY;      PRT;   214 AA.
AC   Q90XQ9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN A.
GN   CLDNA.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359435; AAL01842.1; -.
SQ   SEQUENCE   214 AA;  22952 MW;  8E86F0EB2B72357D CRC64;

  Query Match             36.8%;  Score 432.5;  DB 13;  Length 214;
  Best Local Similarity   38.6%;  Pred. No. 3.7e-30;
  Matches   80;  Conservative   50;  Mismatches   76;  Indels    1;  Gaps    1;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||| |||::|  : ::| :| ::   :| |: ::::| :|| |    :|||| |   |||
Db       1 MASTGLQVLGMAMSIIGWVGCIITCAMPMWRVTAFIGNNIVVAQIIWEGLWMNCIVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             || :| ::| || |:|||:|: |    :: || :: :|| :||   ::   | :|::
Db      61 QMQCKVYDSMLALPQDLQAARALTVICILVALLAMLVGIVGAKCTNCIEDENTKAKVSMV 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
            || |:: |:|  ||| |: : |:||||:||| :: | |:| |:|:|  ||   |: | :
Db     121 SGVVFLVAGILMLIPVCWSANSIIRDFYNPLVVEAQKRELGAAIYIGWASSALMLLGGGL 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRS 207
           || ||  ::| : |   | |     ||
Db     181 LCCSC-PKKNDAPYSARYTAPSGPARS 206


RESULT   8
Q9DE12
ID   Q9DE12      PRELIMINARY;      PRT;   214 AA.
AC   Q9DE12;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TRANSMEMBRANE TIGHT JUNCTION PROTEIN CLAUDIN.
GN   CLA.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=DORSAL LIP;
RX   MEDLINE=21098758; PubMed=11161574;
RA   Brizuela B.J., Wessely O., De Robertis E.M.;
RT   "Overexpression of the Xenopus Tight-Junction Protein Claudin Causes
RT   Randomization of the Left-Right Body Axis.";
RL   Dev. Biol. 230:217-229(2001).
DR   EMBL; AF224712; AAG44257.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; UNKNOWN_1.
SQ   SEQUENCE   214 AA;  23151 MW;  3F0D65E3ADE311E6 CRC64;


  Query Match             36.6%;  Score 430;  DB 13;  Length 214;
  Best Local Similarity   36.5%;  Pred. No. 6.2e-30;
  Matches   84;  Conservative   52;  Mismatches   78;  Indels   16;  Gaps    3;

Qy       1 MASLGLQLVGYILGLLGLLGLTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           ||| |||::|  : ::| :|:::: || |: ::::| :|| |    :|||| |   |||
Db       1 MASTGLQILGMAMSIIGWVGSIISCALPMWRVTAFIGNNIVVAQIIWEGLWMNCIVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             || :| ::| |  |:|||:|: |    :: ||  | :|| :||   ::   | :|::
Db      61 QMQCKVYDSMLALAQDLQAARALTVICILVALLAMFIGIVGAKCTNCIEDENTKAKVSMV 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           |: |:: |:|  ||| |: : |:||||:||| :: | |:| |||:   ||   |: | :
Db     121 SGIVFLVAGILMLIPVCWSANSIIRDFYNPLVVEAQKRELGAALYIRWASSALLLLGGSL 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           || ||  ::| : |   | |          |` ||:     : |:   ||
Db     181 LCCSC-PKKNDAPYPARYTA----------PSCPPR-----SDYTSKNYV 214


RESULT   9
Q98SR2
ID   Q98SR2      PRELIMINARY;      PRT;   214 AA.
AC   Q98SR2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CLAUDIN-3.
GN   CLDN3.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Reardon E., Kojima S., Rizzolo L.J.;
RT   "Sequence of chick claudin-3 cDNA.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF334677; AAK20876.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; UNKNOWN_1.
SQ   SEQUENCE   214 AA;  23090 MW;  A2540116CBD53978 CRC64;


  Query Match             36.5%;  Score 428.5;  DB 13;  Length 214;
  Best Local Similarity   36.4%;  Pred. No. 8.4e-30;
  Matches   83;  Conservative   50;  Mismatches   80;  Indels   15;  Gaps    2;

Qy       3 SLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGIT 62
           |:||:: |  | :|| | :::   || |: ::::| :|||    :|||| |   |||
Db       2 SMGLEIGGVALSVLGWLCSIICCALPMWRVTAFIGNNIVTAQIIWEGLWMNCVVQSTGQM 61

Qy      63 QCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGG 122
             || :| ::| || |:|||:|::| :  :: |  ::::|| :||   ::   | :: :  |
Db      62 QCKVYDSMLALPQDLQAARALLVVAIVLAVLGLMVAIVGAQCTRCVEDETTKAKITIVSG 121

Qy     123 VFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILC 182
           | |:| |::  |||:|: : |:||||:||| |: | |:| :||:|  :|   |  | :||
Db     122 VIFLLSGIMTLIPVSWSANTIIRDFYNPLVIDAQKRELGTSLYVGWAASALLLFGGALLC 181

Qy     183 FSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           ||  :       |   |       |:||        |   ||    ||
Db     182 CSCPPK-------DERYAPSKVAYSAPR--------SAVTSYDKRNYV 214


RESULT  10
Q90XR0
ID   Q90XR0      PRELIMINARY;      PRT;   214 AA.
AC   Q90XR0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN H.
GN   CLDNH.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359434; AAL01841.1; -.
SQ   SEQUENCE   214 AA;  22788 MW;  38163AB7E50BF058 CRC64;


  Query Match             35.9%;  Score 421;  DB 13;  Length 214;
  Best Local Similarity   38.5%;  Pred. No. 3.8e-29;
  Matches   80;  Conservative   49;  Mismatches   65;  Indels   14;  Gaps    1;

Qy       3 SLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGIT 62
           |:||:: |  ||::| : ::||  || |: |::|||:||||    :|||| |   |||
Db       2 SMGLEIGGIALGIIGWIISIVACALPMWRVSAFVGANIVTAQVIWEGLWMNCVVQSTGQM 61

Qy      63 QCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGG 122
           || :| ::| |  |:||::|| | :  :: |  :|||:| :||   ::   || :| :  |
Db      62 QCKVYDSMLALGQDLQASRAMTVIAIILAVLGVMISVMGAKCTNCIEDEGAKAKVMIVSG 121

Qy     123 VFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILC 182
           : ||: |:|  || ||  : |:||||:||:| : : |:| ::|:|  ::   :| | :||
Db     122 IMFIIAGILDLIPSAWVANQIIRDFYNPLLPGAQQRELGASIYIGFAAAALLIIGGAMLC 181

Qy     183 FSC--------------SSQRNRSNYYD 196
           :|               |: |: |  ||
Db     182 CTCPPKEKKYKPARMGYSAPRSASAGYD 209


RESULT  11
Q90XR4
ID   Q90XR4      PRELIMINARY;      PRT;   210 AA.
AC   Q90XR4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN J.
GN   CLDNJ.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359430; AAL01837.1; -.
SQ   SEQUENCE   210 AA;  22895 MW;  F9A6964463E55775 CRC64;


  Query Match             35.6%;  Score 417.5;  DB 13;  Length 210;
  Best Local Similarity   36.1%;  Pred. No. 7.4e-29;
  Matches   82;  Conservative   47;  Mismatches   81;  Indels   17;  Gaps    2;

Qy       4 LGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQ 63
           : ||::|  | ::|  ||::  || || ::::| :|| |  | :|||| |     |  |
Db       1 MALQVLGITLSMIGFAGTIIICALPMWKVTAFIGTNIVVAQVFWEGLWMTCVYERIGQMQ 60

Qy      64 CDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVAGGV 123
           | :|  || |   :||:: ::|||: |::||| :| ::|  ||      ||| |: |  |:
Db      61 CKLYDALLDLDPFLQASRGLIVTTMALASLAFLIFLIGADCTNCLSNPRAKGRIVVVSGI 120

Qy     124 FFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILCF 183
            |:| ||   :||:|    |:|||::|:| :::| |:| |||:| ::: |  : | |||
Db     121 TFMLSGLTTVVPVSWTADSIIRDFHNPVVHEALKREMGAALYVGWLTAGFLFVGGAILCT 180

Qy     184 SCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
           ||  :|:  ||   |     | |                |::  ||
Db     181 SCPPERD--NYLPRYTLTKSGTHS--------------GYAVKNYV 210


RESULT  12
Q90XS1
ID   Q90XS1      PRELIMINARY;      PRT;   206 AA.
AC   Q90XS1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN A.
GN   CLDNA.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359423; AAL01830.1; -.
SQ   SEQUENCE   206 AA;  21694 MW;  69CF7A8554A2EFCF CRC64;


  Query Match             35.2%;  Score 413.5;  DB 13;  Length 206;
  Best Local Similarity   37.4%;  Pred. No. 1.6e-28;
  Matches   77;  Conservative   46;  Mismatches   80;  Indels    3;  Gaps    1;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           | | |||:||  | ::| :| :|  :|| |: ::::| :||||    :|:|| |   |||
Db       1 MVSAGLQMVGIALAVIGWIGVIVVCILPMWQVTAFIGQNIVTAQVTWEGIWMSCVVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             || :| ::| | :|:|||:|: : |  :  :  :::  | :||    | ||| :| |
Db      61 QMQCKVYDSMLALSSDLQAARALCIISILVGVVGILLAAAGGKCTTCIXEERAKAKVCVI 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
             |  ||: |:|  ||| |  : |:|||||:|:   : | |:| :|::|  :|   :| | :
Db     121 SGAIFIVAGVLCLIPVCWTANNIIRDFYNPMTNSAQKRELGASLFIGWGASALLIIGGSL 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATR 206
           || :|  |     |   | |:   |:
Db     181 LCANCPPQ---DQYKATYTARSGGTK 203
```

```
RESULT  13
Q90XR6
ID   Q90XR6      PRELIMINARY;      PRT;   211 AA.
AC   Q90XR6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CLAUDIN I.
GN   CLDNI.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21417738; PubMed=11517306;
RA   Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;
RT   "Expression and phylogeny of claudins in vertebrate primordia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR   EMBL; AF359428; AAL01835.1; -.
SQ   SEQUENCE   211 AA;  22174 MW;  60F9E453891B5639 CRC64;


  Query Match             33.7%;  Score 396;  DB 13;  Length 211;
  Best Local Similarity   38.3%;  Pred. No. 5.6e-27;
  Matches   79;  Conservative   44;  Mismatches   79;  Indels    4;  Gaps    2;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           | | |:|:|   ||:|||: |:| : :| ||||:::| :|:||    :||||:|   |||
Db       1 MGSAGVQIVCVALGILGLIATVVTIAIPQWKTSAFIGQNIITAQVSEEGLWMQCVVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             ||  | :|| | :|:|||:|| : |  :| |: :|   |   |   :    | :| :
Db      61 QQQCKSYDSLLILSSDLQAARAMTIISCMLSVLSLLILCGGADFTTCIENEDVKPKVTLV 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIA-GI 179
             :  || |||  |||:|  : ::||| :|:||::| |:|  :|:|  | :  ::| |:
Db     121 SAIGLILAGLLVIIPVSWAANNVVRDFNNPMVPEAQKRELGPCIYIGWASGVLLILAGGL 180

Qy     180 ILCFS---CSSQRNRSNYYDAYQAQP 202
           : |||    |    : ||    : |
Db     181 LCCFSRPRSSGSSGAAKYYSNSASAP 206


RESULT  14
Q9ET38
ID   Q9ET38      PRELIMINARY;      PRT;   193 AA.
AC   Q9ET38;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CLAUDIN-19 (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RA   Kiuchi Y., Morita K., Furuse M., Tsukita S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF249889; AAF98323.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; UNKNOWN_1.
FT   NON_TER       1      1
FT   NON_TER     193    193
SQ   SEQUENCE   193 AA;  20299 MW;  2F2D82DB5FCF0D7F CRC64;
```

```
  Query Match             33.5%;  Score 393.5;  DB 11;  Length 193;
  Best Local Similarity   39.4%;  Pred. No. 8.4e-27;
  Matches   76;  Conservative   41;  Mismatches   75;  Indels    1;  Gaps    1;

Qy      11 YILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTL 70
           | | | | :| : :  || || ||| | :|:|||| :|||| ||: ||| || :| :|
Db       1 YFLALGGWVGIIASTALPQWKQSSYAGDAIITAVGLYEGLWMSCASQSTGQVQCKLYDSL 60

Qy      71 LGLPADIQAAQAMMVTSSAISSLACIISVVGMRCT-VFCQESRAKDRVAVAGGVFFILGG 129
           | |   ||:|:|:|| :  :  :| ::|||||:|| |       || |||::||  |:| |
Db      61 LALDGHIQSARALMVVAVLLGFVAMVLSVVGMKCTRVGDSNPTAKSRVAISGGALFLLAG 120

Qy     130 LLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGIILCFSCSSQR 189
           |     |:|    : ::|::|  | : ::| | ||::|  |:  ::: |  || :|
Db     121 LCTLTAVSWYATLVTQEFFNPSTPVNARYEFGPALFVGWASAGLAMLGGSFLCCTCPEPE 180

Qy     190 NRSNYYDAYQAQP 202
             ::    |:: |
Db     181 RANSIPQPYRSGP 193


RESULT  15
Q98SR1
ID   Q98SR1      PRELIMINARY;      PRT;   216 AA.
AC   Q98SR1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CLAUDIN-5.
GN   CLDN5.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Reardon E., Kojima S., Rizzolo L.J.;
RT   "Sequence of chick claudin-5 cDNA.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF334678; AAK20877.1; -.
DR   InterPro; IPR001832; Claudin.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
DR   PRINTS; PR01077; CLAUDIN.
DR   PROSITE; PS01346; CLAUDIN; UNKNOWN_1.
SQ   SEQUENCE   216 AA;  23180 MW;  5037A882DFF5433A CRC64;


  Query Match             33.5%;  Score 393;  DB 13;  Length 216;
  Best Local Similarity   36.2%;  Pred. No. 1e-26;
  Matches   75;  Conservative   48;  Mismatches   84;  Indels    0;  Gaps    0;

Qy       1 MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
           |||  ::::|  ||:|| :| ::|  || |: |:::  :|| |    :|||| |   |||
Db       1 MASAAVEILGLGLGILGWVGVILACGLPMWQVSAFIDVNIVVAQTIWEGLWMNCVVQSTG 60

Qy      61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
             || :| ::| |  ::|| :|: |  : :  :| :::||| :||   :  : | |: :|
Db      61 QMQCKVYDSILALRPEVQAGRALTVIVALLGLVALMVTVVGAQCTNCIRPGKMKSRIVIA 120

Qy     121 GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
           ||  :|| |:|  :|: |  : :: ||| | || | | ||| |||:|  ::   |  | :
Db     121 GGTIYILCGVLVLVPLCWFANIVISDFYDPSVPPSQKREIGAALYIGWAATALLLFGGCL 180

Qy     181 LCFSCSSQRNRSNYYDAYQAQPLATRS 207
           :|     ||: :::   | |    | |
Db     181 ICCCSCLQRDETSFPVKYSAPRRPTSS 207


Search completed: June  2, 2002, 20:15:55

Job time: 938 sec
```